(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
23 January 2003 (23.01.2003)

PCT

(10) International Publication Number
WO 03/006618 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/21971

(22) International Filing Date: 10 July 2002 (10.07.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/305,089	12 July 2001 (12.07.2001)	US
60/305,104	12 July 2001 (12.07.2001)	US
60/305,325	13 July 2001 (13.07.2001)	US
60/305,390	13 July 2001 (13.07.2001)	US
60/306,960	19 July 2001 (19.07.2001)	US
60/306,694	20 July 2001 (20.07.2001)	US
60/308,170	27 July 2001 (27.07.2001)	US

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **SWARNAKAR, Anita** [CA/US]; 8 Locksley Avenue #5D, San Francisco, CA 94122 (US). **RICHARDSON, Thomas, W.** [US/US]; 616 Canyon Road #107, Redwood City, CA 94062 (US). **WARREN, Bridget, A.** [US/US]; 1810 S. El Camino Real #B103, Encitas, CA 92024 (US). **GRIFFIN, Jennifer, A.** [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **EMERLING, Brooke, M.** [US/US]; 1735 Woodland Avenue # 71, Palo Alto, CA 94303 (US). **LAL, Preeti, G.** [US/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **FORSYTHE, Ian, J.** [US/US]; 308 Roble Avenue, Redwood City, CA 94061 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **LI, Joana, X.** [US/US]; 1264 Geneva Avenue, San Francisco, CA 94112 (US). **BECHA, Shanya, D.** [US/US]; 21062 Gary Drive # 117, Castro Valley, CA 94546 (US). **DUGGAN, Brendan, M.** [AU/US]; 243 Buena Vista Avenue # 306, Sunnyvale, CA 94086 (US). **SANJANWALA, Madhusudan, M.** [US/US]; 210 Silvia Court, Los Altos, CA 94024 (US). **LEE, Ernestine, A.** [US/US]; 20523 Crow Creek Road, Castro Valley, CA 94552 (US). **BURFORD, Neil** [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). **ELLIOTT, Vicki, S.** [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). **ISON, Craig, H.** [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US). **DING, Li** [CN/US]; 11036 Ridgecorde Lane, Creve Coeur, MO 63141 (US). **BOROWSKY, Mark, L.** [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US). **YAO, Monique, G.** [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). **BARROSO, Ines** [PT/GB]; 38 Eden Street, Cambridge CB1 1EL (GB). **TRAN, Bao** [US/US]; 750 Salberg Avenue, Santa Clara, CA 95051 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street A # 205, San Leandro, CA 94577 (US). **HAFALIA, April, J., A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **NGUYEN, Danniel, B.** [US/US]; 4922 Arnica Court, San Jose, CA 95111 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). **ARVIZU, Chandra, S.** [US/US]; 1706 Morocco Drive, San Jose, CA 95125 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NUCLEIC ACID-ASSOCIATED PROTEINS

(57) Abstract: Various embodiments of the invention provide human nucleic acid-associated proteins (NAAP) and polynucleotides which identify and encode NAAP. Embodiments of the invention also provide expression vectors, host cells, antibodies, agonists, and antagonists. Other embodiments provide methods for diagnosing, treating, or preventing disorders associated with aberrant expression of NAAP.

WO 03/006618 A2

NUCLEIC ACID-ASSOCIATED PROTEINS

TECHNICAL FIELD

The invention relates to novel nucleic acids, nucleic acid-associated proteins encoded by these nucleic acids, and to the use of these nucleic acids and proteins in the diagnosis, treatment, and prevention of cell proliferative, neurological, developmental, and autoimmune/inflammatory disorders, and infections. The invention also relates to the assessment of the effects of exogenous compounds on the expression of nucleic acids and nucleic acid-associated proteins.

BACKGROUND OF THE INVENTION

Multicellular organisms are comprised of diverse cell types that differ dramatically both in structure and function. The identity of a cell is determined by its characteristic pattern of gene expression, and different cell types express overlapping but distinctive sets of genes throughout development. Spatial and temporal regulation of gene expression is critical for the control of cell proliferation, cell differentiation, apoptosis, and other processes that contribute to organismal development. Furthermore, gene expression is regulated in response to extracellular signals that mediate cell-cell communication and coordinate the activities of different cell types. Appropriate gene regulation also ensures that cells function efficiently by expressing only those genes whose functions are required at a given time.

Transcription Factors

Transcriptional regulatory proteins are essential for the control of gene expression. Some of these proteins function as transcription factors that initiate, activate, repress, or terminate gene transcription. Transcription factors generally bind to the promoter, enhancer, and upstream regulatory regions of a gene in a sequence-specific manner, although some factors bind regulatory elements within or downstream of a gene coding region. Transcription factors may bind to a specific region of DNA singly or as a complex with other accessory factors. (Reviewed in Lewin, B. (1990) Genes IV, Oxford University Press, New York, NY, and Cell Press, Cambridge, MA, pp. 554-570.)

The double helix structure and repeated sequences of DNA create topological and chemical features which can be recognized by transcription factors. These features are hydrogen bond donor and acceptor groups, hydrophobic patches, major and minor grooves, and regular, repeated stretches of sequence which induce distinct bends in the helix. Typically, transcription factors recognize specific DNA sequence motifs of about 20 nucleotides in length. Multiple, adjacent transcription factor-binding motifs may be required for gene regulation.

Many transcription factors incorporate DNA-binding structural motifs which comprise either a helices or ß sheets that bind to the major groove of DNA. Four well-characterized structural motifs

are helix-turn-helix, zinc finger, leucine zipper, and helix-loop-helix. Proteins containing these motifs may act alone as monomers, or they may form homo- or heterodimers that interact with DNA.

The helix-turn-helix motif consists of two a helices connected at a fixed angle by a short chain of amino acids. One of the helices binds to the major groove. Helix-turn-helix motifs are exemplified by the homeobox motif which is present in homeodomain proteins. These proteins are critical for specifying the anterior-posterior body axis during development and are conserved throughout the animal kingdom. The Antennapedia and Ultrabithorax proteins of *Drosophila melanogaster* are prototypical homeodomain proteins. (Pabo, C.O. and R.T. Sauer (1992) Annu. Rev. Biochem. 61:1053-1095.)

The zinc finger motif, which binds zinc ions, generally contains tandem repeats of about 30 amino acids consisting of periodically spaced cysteine and histidine residues. Examples of this sequence pattern, designated C2H2 and C3HC4 ("RING" finger), have been described. (Lewin, *supra.*) Zinc finger proteins each contain an a helix and an antiparallel ß sheet whose proximity and conformation are maintained by the zinc ion. Contact with DNA is made by the arginine preceding the a helix and by the second, third, and sixth residues of the a helix. Variants of the zinc finger motif include poorly defined cysteine-rich motifs which bind zinc or other metal ions. These motifs may not contain histidine residues and are generally nonrepetitive. The zinc finger motif may be repeated in a tandem array within a protein, such that the a helix of each zinc finger in the protein makes contact with the major groove of the DNA double helix. This repeated contact between the protein and the DNA produces a strong and specific DNA-protein interaction. The strength and specificity of the interaction can be regulated by the number of zinc finger motifs within the protein. Though originally identified in DNA-binding proteins as regions that interact directly with DNA, zinc fingers occur in a variety of proteins that do not bind DNA (Lodish, H. et al. (1995) Molecular Cell Biology, Scientific American Books, New York, NY, pp. 447-451). For example, Galcheva-Gargova, Z. et al. ((1996) Science 272:1797-1802) have identified zinc finger proteins that interact with various cytokine receptors.

The C2H2-type zinc finger signature motif contains a 28 amino acid sequence, including 2 conserved Cys and 2 conserved His residues in a C-2-C-12-H-3-H type motif. The motif generally occurs in multiple tandem repeats. A cysteine-rich domain including the motif Asp-His-His-Cys (DHHC-CRD) has been identified as a distinct subgroup of zinc finger proteins. The DHHC-CRD region has been implicated in growth and development. One DHHC-CRD mutant shows defective function of Ras, a small membrane-associated GTP-binding protein that regulates cell growth and differentiation, while other DHHC-CRD proteins probably function in pathways not involving Ras (Bartels, D.J. et al. (1999) Mol. Cell Biol. 19:6775-6787).

Zinc-finger transcription factors are often accompanied by modular sequence motifs such as

the Kruppel-associated box (KRAB) and the SCAN domain. For example, the hypoalphalipoproteinemia susceptibility gene ZNF202 encodes a SCAN box and a KRAB domain followed by eight C2H2 zinc-finger motifs (Honer, C. et al. (2001) Biochim. Biophys. Acta 1517:441-448). The SCAN domain is a highly conserved, leucine-rich motif of approximately 60 amino acids found at the amino-terminal end of zinc finger transcription factors. SCAN domains are most often linked to C2H2 zinc finger motifs through their carboxyl-terminal end. Biochemical binding studies have established the SCAN domain as a selective hetero- and homotypic oligomerization domain. SCAN domain-mediated protein complexes may function to modulate the biological function of transcription factors (Schumacher, C. et al. (2000) J. Biol. Chem. 275:17173-17179).

The KRAB (Kruppel-associated box) domain is a conserved amino acid sequence spanning approximately 75 amino acids and is found in almost one-third of the 300 to 700 genes encoding C2H2 zinc fingers. The KRAB domain is found N-terminally with respect to the finger repeats. The KRAB domain is generally encoded by two exons; the KRAB-A region or box is encoded by one exon and the KRAB-B region or box is encoded by a second exon. The function of the KRAB domain is the repression of transcription. Transcription repression is accomplished by recruitment of either the KRAB-associated protein-1, a transcriptional corepressor, or the KRAB-A interacting protein. Proteins containing the KRAB domain are likely to play a regulatory role during development (Williams, A.J. et al. (1999) Mol. Cell Biol. 19:8526-8535). A subgroup of highly related human KRAB zinc finger proteins detectable in all human tissues is highly expressed in human T lymphoid cells (Bellefroid, E.J. et al. (1993) EMBO J. 12:1363-1374). The ZNF85 KRAB zinc finger gene, a member of the human ZNF91 family, is highly expressed in normal adult testis, in seminomas, and in the NT2/D1 teratocarcinoma cell line (Poncelet, D.A. et al. (1998) DNA Cell Biol.17:931-943).

Additional zinc finger-associated proteins include the sprouty (SPRY) protein, which was first identified in a genetic screen in *Drosophila*. SPRY proteins are classified by virtue of their characteristic cysteine-rich residues located in their carboxyl termini (Wong, E.S.M., et al. (2001) J. Biol. Chem. 276:5866-5875). Zinc-binding B-box motifs are located within the B30.2-like domain, constituting a diverse family of proteins (Seto, M.H., et al. (1999) Proteins 35:235-249). The functions of these domains include regulation of cell growth and differentiation. The SPRY domain has been identified as a subdomain within the B30.2-like domain (Torok, M. and Etkin, L.D. (2001) Differentiation 67:63-71). The B-box domain itself is involved in growth control and transcriptional regulation. These genes possess several conserved motifs that always include a B-box zinc binding motif associated with various other motifs such as the RING zinc finger. The RING finger domain is a zinc-binding Cys-His protein motif found in various proteins involved in signal transduction, gene

transcription, differentiation, and morphogenesis. A RING-B-box-coiled-coil (RBCC) subclass of RING-finger proteins contains an NH_2-terminal RING-finger followed by either single or multiple additional B-box zinc finger domains (Spencer, J.A., et al. (2000) J. Cell Biol. 150:771-784). Several RBCC proteins have been implicated in oncogenesis. The RET finger protein (RFP) also belongs to the B-box zinc finger protein family. RFPs possess a tripartite motif consisting of a RING finger, a B-box finger, and a coiled-coil domain. RFP may become oncogenic when its tripartite motif becomes fused with the tyrosine kinase domain of the RET protein (Tezel, G., et al. (1999) Pathol. Int. 49:881-886).

The C4 motif is found in hormone-regulated proteins. The C4 motif generally includes only 2 repeats. A number of eukaryotic and viral proteins contain a conserved cysteine-rich domain of 40 to 60 residues (called C3HC4 zinc-finger or RING finger) that binds two atoms of zinc, and is probably involved in mediating protein-protein interactions. The 3D "cross-brace" structure of the zinc ligation system is unique to the RING domain. The spacing of the cysteines in such a domain is C-x(2)-C-x(9 to 39)-C-x(1 to 3)-H-x(2 to3)-C-x(2)-C-x(4 to 48)-C-x(2)-C. The PHD finger is a C4HC3 zinc-finger-like motif found in nuclear proteins thought to be involved in chromatin-mediated transcriptional regulation.

GATA-type transcription factors contain one or two zinc finger domains which bind specifically to a region of DNA that contains the consecutive nucleotide sequence GATA. NMR studies indicate that the zinc finger comprises two irregular anti-parallel b sheets and an a helix, followed by a long loop to the C-terminal end of the finger (Ominchinski, J.G. (1993) Science 261:438-446). The helix and the loop connecting the two b-sheets contact the major groove of the DNA, while the C-terminal part, which determines the specificity of binding, wraps around into the minor groove.

The LIM motif consists of about 60 amino acid residues and contains seven conserved cysteine residues and a histidine within a consensus sequence (Schmeichel, K.L. and M.C. Beckerle (1994) Cell 79:211-219). The LIM family includes transcription factors and cytoskeletal proteins which may be involved in development, differentiation, and cell growth. One example is actin-binding LIM protein, which may play roles in regulation of the cytoskeleton and cellular morphogenesis (Roof, D.J. et al. (1997) J. Cell Biol. 138:575-588). The N-terminal domain of actin-binding LIM protein has four double zinc finger motifs with the LIM consensus sequence. The C-terminal domain of actin-binding LIM protein shows sequence similarity to known actin-binding proteins such as dematin and villin. Actin-binding LIM protein binds to F-actin through its dematin-like C-terminal domain. The LIM domain may mediate protein-protein interactions with other LIM-binding proteins.

Myeloid cell development is controlled by tissue-specific transcription factors. Myeloid zinc

finger proteins (MZF) include MZF-1 and MZF-2. MZF-1 functions in regulation of the development of neutrophilic granulocytes. A murine homolog MZF-2 is expressed in myeloid cells, particularly in the cells committed to the neutrophilic lineage. MZF-2 is down-regulated by G-CSF and appears to have a unique function in neutrophil development (Murai, K. et al. (1997) Genes Cells 2:581-591).

The leucine zipper motif comprises a stretch of amino acids rich in leucine which can form an amphipathic a helix. This structure provides the basis for dimerization of two leucine zipper proteins. The region adjacent to the leucine zipper is usually basic, and upon protein dimerization, is optimally positioned for binding to the major groove. Proteins containing such motifs are generally referred to as bZIP transcription factors. The leucine zipper motif is found in the proto-oncogenes Fos and Jun, which comprise the heterodimeric transcription factor AP1 involved in cell growth and the determination of cell lineage (Papavassiliou, A.G. (1995) N. Engl. J. Med. 332:45-47).

Many neoplastic disorders in humans can be attributed to inappropriate gene expression. Malignant cell growth may result from either excessive expression of tumor promoting genes or insufficient expression of tumor suppressor genes (Cleary, M.L. (1992) Cancer Surv. 15:89-104). Chromosomal translocations may also produce chimeric loci which fuse the coding sequence of one gene with the regulatory regions of a second unrelated gene. Such an arrangement likely results in inappropriate gene transcription, potentially contributing to malignancy. One clinically relevant zinc-finger protein is WT1, a tumor-suppressor protein that is inactivated in children with Wilm's tumor. The oncogene bcl-6, which plays an important role in large-cell lymphoma, is also a zinc-finger protein (Papavassiliou, A.G. *supra*).

The helix-loop-helix motif (HLH) consists of a short a helix connected by a loop to a longer a helix. The loop is flexible and allows the two helices to fold back against each other and to bind to DNA. The transcription factor Myc contains a prototypical HLH motif.

The NF-kappa-B/Rel signature defines a family of eukaryotic transcription factors involved in oncogenesis, embryonic development, differentiation and immune response. Most transcription factors containing the Rel homology domain (RHD) bind as dimers to a consensus DNA sequence motif termed kappa-B. Members of the Rel family share a highly conserved 300 amino acid domain termed the Rel homology domain. The characteristic Rel C-terminal domain is involved in gene activation and cytoplasmic anchoring functions. Proteins known to contain the RHD domain include vertebrate nuclear factor NF-kappa-B, which is a heterodimer of a DNA-binding subunit and the transcription factor p65, mammalian transcription factor RelB, and vertebrate proto-oncogene c-rel, a protein associated with differentiation and lymphopoiesis (Kabrun, N. and P.J. Enrietto (1994) Semin. Cancer Biol. 5:103-112).

A DNA binding motif termed ARID (AT-rich interactive domain) distinguishes an evolutionarily conserved family of proteins. The approximately 100-residue ARID sequence is

present in a series of proteins strongly implicated in the regulation of cell growth, development, and tissue-specific gene expression. ARID proteins include Bright (a regulator of B-cell-specific gene expression), dead ringer (involved in development), and MRF-2 (which represses expression from the cytomegalovirus enhancer) (Dallas, P.B. et al. (2000) Mol. Cell Biol. 20:3137-3146).

The ELM2 (Egl-27 and MTA1 homology 2) domain is found in metastasis-associated protein MTA1 and protein ER1. The *Caenorhabditis elegans* gene egl-27 is required for embryonic patterning MTA1, a human gene with elevated expression in metastatic carcinomas, is a component of a protein complex with histone deacetylase and nucleosome remodelling activities (Solari, F. et al. (1999) Development 126:2483-2494). The ELM2 domain is usually found to the N terminus of a myb-like DNA binding domain. ELM2 is also found associated with an ARID DNA.

The Iroquois (Irx) family of genes are found in nematodes, insects and vertebrates. Irx genes usually occur in one or two genomic clusters of three genes each and encode transcriptional controllers that possess a characteristic homeodomain. The Irx genes function early in development to specify the identity of diverse territories of the body. Later in development in both *Drosophila* and vertebrates, the Irx genes function again to subdivide those territories into smaller domains. (For a review of Iroquois genes, see Cavodeassi, F. et al. (2001) Development 128:2847-2855.) For example, mouse and human Irx4 proteins are 83% conserved and their 63-aa homeodomain is more than 93% identical to that of the *Drosophila* Iroquois patterning genes. Irx4 transcripts are predominantly expressed in the cardiac ventricles. The homeobox gene Irx4 mediates ventricular differentiation during cardiac development (Bruneau, B.G. et al. (2000) Dev. Biol. 217:266-77).

Histidine triad (HIT) proteins share residues in distinctive dimeric, 10-stranded half-barrel structures that form two identical purine nucleotide-binding sites. Hint (histidine triad nucleotide-binding protein)-related proteins, found in all forms of life, and fragile histidine triad (Fhit)-related proteins, found in animals and fungi, represent the two main branches of the HIT superfamily. Fhit homologs bind and cleave diadenosine polyphosphates. Fhit-Ap(n)A complexes appear to function in a proapoptotic tumor suppression pathway in epithelial tissues (Brenner C. et al. (1999) J. Cell Physiol.181:179-187).

Most transcription factors contain characteristic DNA binding motifs, and variations on the above motifs and new motifs have been and are currently being characterized. (Faisst, S. and S. Meyer (1992) Nucleic Acids Res. 20:3-26.)

Most transcription factors contain characteristic DNA binding motifs, and variations on the above motifs and new motifs have been and are currently being characterized (Faisst, S. and S. Meyer (1992) Nucl. Acids Res. 20:3-26). These include the forkhead motif, found in transcription factors involved in development and oncogenesis (Hacker, U. et al. (1995) EMBO J. 14:5306-5317). Foxj2 is a human forkhead transcriptional activator that binds DNA with a dual sequence specificity. Foxj2

expression is activated early in zygotic development (Granadino, B. et al. (2000) Mech. Dev. 97:157-160).

Cold-shock proteins (Csp) are involved in a specific pattern of gene expression in response to abrupt shifts to lower temperatures. This pattern includes the induction of cold-shock proteins, synthesis of proteins involved in transcription and translation, and repression of heat-shock proteins. The major cold-shock protein, cold-shock protein A (CspA), has high sequence similarity with three other proteins--CspB, CspC, and CspD. The Csp proteins share sequence similarity with other prokaryotic proteins and with the 'cold-shock domain' of eukaryotic Y-box proteins (Jones, P.G. and Inouye, M. (1994) Mol. Microbiol. 11:811-818).

Chromatin Associated Proteins

In the nucleus, DNA is packaged into chromatin, the compact organization of which limits the accessibility of DNA to transcription factors and plays a key role in gene regulation. (Lewin, *supra*, pp. 409-410.) The compact structure of chromatin is determined and influenced by chromatin-associated proteins such as the histones, the high mobility group (HMG) proteins, and the chromodomain proteins. There are five classes of histones, H1, H2A, H2B, H3, and H4, all of which are highly basic, low molecular weight proteins. The fundamental unit of chromatin, the nucleosome, consists of 200 base pairs of DNA associated with two copies each of H2A, H2B, H3, and H4. H1 links adjacent nucleosomes. HMG proteins are low molecular weight, non-histone proteins that may play a role in unwinding DNA and stabilizing single-stranded DNA. Chromodomain proteins play a key role in the formation of highly compacted heterochromatin, which is transcriptionally silent.

Diseases and Disorders Related to Gene Regulation

Many neoplastic disorders in humans can be attributed to inappropriate gene expression. Malignant cell growth may result from either excessive expression of tumor promoting genes or insufficient expression of tumor suppressor genes (Cleary, M.L. (1992) Cancer Surv. 15:89-104). The zinc finger-type transcriptional regulator WT1 is a tumor-suppressor protein that is inactivated in children with Wilm's tumor. The oncogene bcl-6, which plays an important role in large-cell lymphoma, is also a zinc-finger protein (Papavassiliou, A.G. (1995) N. Engl. J. Med. 332:45-47). Chromosomal translocations may also produce chimeric loci that fuse the coding sequence of one gene with the regulatory regions of a second unrelated gene. Such an arrangement likely results in inappropriate gene transcription, potentially contributing to malignancy. In Burkitt's lymphoma, for example, the transcription factor Myc is translocated to the immunoglobulin heavy chain locus, greatly enhancing Myc expression and resulting in rapid cell growth leading to leukemia (Latchman, D.S. (1996) N. Engl. J. Med. 334:28-33).

In addition, the immune system responds to infection or trauma by activating a cascade of events that coordinate the progressive selection, amplification, and mobilization of cellular defense

mechanisms. A complex and balanced program of gene activation and repression is involved in this process. However, hyperactivity of the immune system as a result of improper or insufficient regulation of gene expression may result in considerable tissue or organ damage. This damage is well-documented in immunological responses associated with arthritis, allergens, heart attack, stroke, and infections (Isselbacher, K.J. et al. Harrison's Principles of Internal Medicine, 13/e, McGraw Hill, Inc. and Teton Data Systems Software, 1996). The causative gene for autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) was recently isolated and found to encode a protein with two PHD-type zinc finger motifs (Bjorses, P. et al. (1998) Hum. Mol. Genet. 7:1547-1553).

Furthermore, the generation of multicellular organisms is based upon the induction and coordination of cell differentiation at the appropriate stages of development. Central to this process is differential gene expression, which confers the distinct identities of cells and tissues throughout the body. Failure to regulate gene expression during development could result in developmental disorders. Human developmental disorders caused by mutations in zinc finger-type transcriptional regulators include: urogenital developmental abnormalities associated with WT1; Greig cephalopolysyndactyly, Pallister-Hall syndrome, and postaxial polydactyly type A (GLI3), and Townes-Brocks syndrome, characterized by anal, renal, limb, and ear abnormalities (SALL1) (Engelkamp, D. and V. van Heyningen (1996) Curr. Opin. Genet. Dev. 6:334-342; Kohlhase, J. et al. (1999) Am. J. Hum. Genet. 64:435-445).

Human acute leukemias involve reciprocal chromosome translocations that fuse the ALL-1 gene located at chromosome region 11q23 to a series of partner genes positioned on a variety of human chromosomes. The fused genes encode chimeric proteins. The AF17 gene encodes a protein of 1093 amino acids, containing a leucine-zipper dimerization motif located 3' of the fusion point and a cysteine-rich domain at the N terminus that shows homology to a domain within the protein Br140 (peregrin) (Prasad R. et al. (1994) Proc. Natl. Acad. Sci. USA 91:8107-8111).

SYNTHESIS OF NUCLEIC ACIDS

Polymerases

DNA and RNA replication are critical processes for cell replication and function. DNA and RNA replication are mediated by the enzymes DNA and RNA polymerase, respectively, by a "templating" process in which the nucleotide sequence of a DNA or RNA strand is copied by complementary base-pairing into a complementary nucleic acid sequence of either DNA or RNA. However, there are fundamental differences between the two processes.

DNA polymerase catalyzes the stepwise addition of a deoxyribonucleotide to the 3'-OH end of a polynucleotide strand (the primer strand) that is paired to a second (template) strand. The new DNA strand therefore grows in the 5' to 3' direction (Alberts, B. et al. (1994) The Molecular Biology

of the Cell, Garland Publishing Inc., New York, NY, pp 251-254). The substrates for the polymerization reaction are the corresponding deoxynucleotide triphosphates which must base-pair with the correct nucleotide on the template strand in order to be recognized by the polymerase. Because DNA exists as a double-stranded helix, each of the two strands may serve as a template for the formation of a new complementary strand. Each of the two daughter cells of a dividing cell therefore inherits a new DNA double helix containing one old and one new strand. Thus, DNA is said to be replicated "semiconservatively" by DNA polymerase. In addition to the synthesis of new DNA, DNA polymerase is also involved in the repair of damaged DNA as discussed below under "Ligases."

In contrast to DNA polymerase, RNA polymerase uses a DNA template strand to "transcribe" DNA into RNA using ribonucleotide triphosphates as substrates. Like DNA polymerization, RNA polymerization proceeds in a 5' to 3' direction by addition of a ribonucleoside monophosphate to the 3'-OH end of a growing RNA chain. DNA transcription generates messenger RNAs (mRNA) that carry information for protein synthesis, as well as the transfer, ribosomal, and other RNAs that have structural or catalytic functions. In eukaryotes, three discrete RNA polymerases synthesize the three different types of RNA (Alberts, *supra*, pp. 367-368). RNA polymerase I makes the large ribosomal RNAs, RNA polymerase II makes the mRNAs that will be translated into proteins, and RNA polymerase III makes a variety of small, stable RNAs, including 5S ribosomal RNA and the transfer RNAs (tRNA). In all cases, RNA synthesis is initiated by binding of the RNA polymerase to a promoter region on the DNA and synthesis begins at a start site within the promoter. Synthesis is completed at a stop (termination) signal in the DNA whereupon both the polymerase and the completed RNA chain are released.

Ligases

DNA repair is the process by which accidental base changes, such as those produced by oxidative damage, hydrolytic attack, or uncontrolled methylation of DNA, are corrected before replication or transcription of the DNA can occur. Because of the efficiency of the DNA repair process, fewer than one in a thousand accidental base changes causes a mutation (Alberts, *supra*, pp. 245-249). The three steps common to most types of DNA repair are (1) excision of the damaged or altered base or nucleotide by DNA nucleases, (2) insertion of the correct nucleotide in the gap left by the excised nucleotide by DNA polymerase using the complementary strand as the template and, (3) sealing the break left between the inserted nucleotide(s) and the existing DNA strand by DNA ligase. In the last reaction, DNA ligase uses the energy from ATP hydrolysis to activate the 5' end of the broken phosphodiester bond before forming the new bond with the 3'-OH of the DNA strand. In Bloom's syndrome, an inherited human disease, individuals are partially deficient in DNA ligation and consequently have an increased incidence of cancer (Alberts, *supra* p. 247).

Nucleases

Nucleases comprise enzymes that hydrolyze both DNA (DNase) and RNA (Rnase). They serve different purposes in nucleic acid metabolism. Nucleases hydrolyze the phosphodiester bonds between adjacent nucleotides either at internal positions (endonucleases) or at the terminal 3' or 5' nucleotide positions (exonucleases). A DNA exonuclease activity in DNA polymerase, for example, serves to remove improperly paired nucleotides attached to the 3'-OH end of the growing DNA strand by the polymerase and thereby serves a "proofreading" function. As mentioned above, DNA endonuclease activity is involved in the excision step of the DNA repair process.

RNases also serve a variety of functions. For example, RNase P is a ribonucleoprotein enzyme which cleaves the 5' end of pre-tRNAs as part of their maturation process. RNase H digests the RNA strand of an RNA/DNA hybrid. Such hybrids occur in cells invaded by retroviruses, and RNase H is an important enzyme in the retroviral replication cycle. Pancreatic RNase secreted by the pancreas into the intestine hydrolyzes RNA present in ingested foods. RNase activity in serum and cell extracts is elevated in a variety of cancers and infectious diseases (Schein, C.H. (1997) Nat. Biotechnol. 15:529-536). Regulation of RNase activity is being investigated as a means to control tumor angiogenesis, allergic reactions, viral infection and replication, and fungal infections.

MODIFICATION OF NUCLEIC ACIDS

Methylases

Methylation of specific nucleotides occurs in both DNA and RNA, and serves different functions in the two macromolecules. Methylation of cytosine residues to form 5-methyl cytosine in DNA occurs specifically in CG sequences which are base-paired with one another in the DNA double-helix. The pattern of methylation is passed from generation to generation during DNA replication by an enzyme called "maintenance methylase" that acts preferentially on those CG sequences that are base-paired with a CG sequence that is already methylated. Such methylation appears to distinguish active from inactive genes by preventing the binding of regulatory proteins that "turn on" the gene, but permiting the binding of proteins that inactivate the gene (Alberts, *supra* pp. 448-451). In RNA metabolism, "tRNA methylase" produces one of several nucleotide modifications in tRNA that affect the conformation and base-pairing of the molecule and facilitate the recognition of the appropriate mRNA codons by specific tRNAs. The primary methylation pattern is the dimethylation of guanine residues to form N,N-dimethyl guanine.

Helicases and Single-stranded Binding Proteins

Helicases are enzymes that destabilize and unwind double helix structures in both DNA and RNA. Since DNA replication occurs more or less simultaneously on both strands, the two strands must first separate to generate a replication "fork" for DNA polymerase to act on. Two types of replication proteins contribute to this process, DNA helicases and single-stranded binding proteins.

DNA helicases hydrolyze ATP and use the energy of hydrolysis to separate the DNA strands. Single-stranded binding proteins (SSBs) then bind to the exposed DNA strands, without covering the bases, thereby temporarily stabilizing them for templating by the DNA polymerase (Alberts, *supra*, pp. 255-256).

RNA helicases also alter and regulate RNA conformation and secondary structure. Like the DNA helicases, RNA helicases utilize energy derived from ATP hydrolysis to destabilize and unwind RNA duplexes. The most well-characterized and ubiquitous family of RNA helicases is the DEAD-box family, so named for the conserved B-type ATP-binding motif which is diagnostic of proteins in this family. Over 40 DEAD-box helicases have been identified in organisms as diverse as bacteria, insects, yeast, amphibians, mammals, and plants. DEAD-box helicases function in diverse processes such as translation initiation, splicing, ribosome assembly, and RNA editing, transport, and stability. Examples of these RNA helicases include yeast Drs1 protein, which is involved in ribosomal RNA processing; yeast TIF1 and TIF2 and mammalian eIF-4A, which are essential to the initiation of RNA translation; and human p68 antigen, which regulates cell growth and division (Ripmaster, T.L. et al. (1992) Proc. Natl. Acad. Sci. USA 89:11131-11135; Chang, T.-H. et al. (1990) Proc. Natl. Acad. Sci. USA 87:1571-1575). These RNA helicases demonstrate strong sequence homology over a stretch of some 420 amino acids. Included among these conserved sequences are the consensus sequence for the A motif of an ATP binding protein; the "DEAD box" sequence, associated with ATPase activity; the sequence SAT, associated with the actual helicase unwinding region; and an octapeptide consensus sequence, required for RNA binding and ATP hydrolysis (Pause, A. et al. (1993) Mol. Cell Biol. 13:6789-6798). Differences outside of these conserved regions are believed to reflect differences in the functional roles of individual proteins (Chang et al., *supra*).

Some DEAD-box helicases play tissue- and stage-specific roles in spermatogenesis and embryogenesis. Overexpression of the DEAD-box 1 protein (DDX1) may play a role in the progression of neuroblastoma (Nb) and retinoblastoma (Rb) tumors (Godbout, R. et al. (1998) J. Biol. Chem. 273:21161-21168). These observations suggest that DDX1 may promote or enhance tumor progression by altering the normal secondary structure and expression levels of RNA in cancer cells. Other DEAD-box helicases have been implicated either directly or indirectly in tumorigenesis. (Discussed in Godbout et al., *supra*.) For example, murine p68 is mutated in ultraviolet light-induced tumors, and human DDX6 is located at a chromosomal breakpoint associated with B-cell lymphoma. Similarly, a chimeric protein comprised of DDX10 and NUP98, a nucleoporin protein, may be involved in the pathogenesis of certain myeloid malignancies.

Topoisomerases

Besides the need to separate DNA strands prior to replication, the two strands must be "unwound" from one another prior to their separation by DNA helicases. This function is performed

by proteins known as DNA topoisomerases. DNA topoisomerase effectively acts as a reversible nuclease that hydrolyzes a phosphodiesterase bond in a DNA strand, permits the two strands to rotate freely about one another to remove the strain of the helix, and then rejoins the original phosphodiester bond between the two strands. Topoisomerases are essential enzymes responsible for the topological rearrangement of DNA brought about by transcription, replication, chromatin formation, recombination, and chromosome segregation. Superhelical coils are introduced into DNA by the passage of processive enzymes such as RNA polymerase, or by the separation of DNA strands by a helicase prior to replication. Knotting and concatenation can occur in the process of DNA synthesis, storage, and repair. All topoisomerases work by breaking a phosphodiester bond in the ribose-phosphate backbone of DNA. A catalytic tyrosine residue on the enzyme makes a nucleophilic attack on the scissile phosphodiester bond, resulting in a reaction intermediate in which a covalent bond is formed between the enzyme and one end of the broken strand. A tyrosine-DNA phosphodiesterase functions in DNA repair by hydrolyzing this bond in occasional dead-end topoisomerase I-DNA intermediates (Pouliot, J.J. et al. (1999) Science 286:552-555).

Two types of DNA topoisomerase exist, types I and II. Type I topoisomerases work as monomers, making a break in a single strand of DNA while type II topoisomerases, working as homodimers, cleave both strands. DNA Topoisomerase I causes a single-strand break in a DNA helix to allow the rotation of the two strands of the helix about the remaining phosphodiester bond in the opposite strand. DNA topoisomerase II causes a transient break in both strands of a DNA helix where two double helices cross over one another. This type of topoisomerase can efficiently separate two interlocked DNA circles (Alberts, *supra*, pp.260-262). Type II topoisomerases are largely confined to proliferating cells in eukaryotes, such as cancer cells. For this reason they are targets for anticancer drugs. Topoisomerase II has been implicated in multi-drug resistance (MDR) as it appears to aid in the repair of DNA damage inflicted by DNA binding agents such as doxorubicin and vincristine (DNA topoisomerases are reviewed in Wang, J.C. (1996) Annu. Rev. Biochem. 65:635-692.).

The topoisomerase I family includes topoisomerases I and III (topo I and topo III). The crystal structure of human topoisomerase I suggests that rotation about the intact DNA strand is partially controlled by the enzyme. In this "controlled rotation" model, protein-DNA interactions limit the rotation, which is driven by torsional strain in the DNA (Stewart, L. et al. (1998) Science 379:1534-1541). Structurally, topo I can be recognized by its catalytic tyrosine residue and a number of other conserved residues in the active site region. Topo I is thought to function during transcription. Two topo IIIs are known in humans, and they are homologous to prokaryotic topoisomerase I, with a conserved tyrosine and active site signature specific to this family. Topo III has been suggested to play a role in meiotic recombination. A mouse topo III is highly expressed in

testis tissue and its expression increases with the increase in the number of cells in pachytene (Seki, T. et al. (1998) J. Biol. Chem. 273:28553-28556).

The topoisomerase II family includes two isozymes (IIa and IIb) encoded by different genes. Topo II cleaves double stranded DNA in a reproducible, nonrandom fashion, preferentially in an AT rich region, but the basis of cleavage site selectivity is not known. Structurally, topo II is made up of four domains, the first two of which are structurally similar and probably distantly homologous to similar domains in eukaryotic topo I. The second domain bears the catalytic tyrosine, as well as a highly conserved pentapeptide. The IIa isoform appears to be responsible for unlinking DNA during chromosome segregation. Cell lines expressing IIa but not IIb suggest that IIb is dispensable in cellular processes; however, IIb knockout mice died perinatally due to a failure in neural development. That the major abnormalities occurred in predominantly late developmental events (neurogenesis) suggests that IIb is needed not at mitosis, but rather during DNA repair (Yang, X. et al. (2000) Science 287:131-134).

Topoisomerases have been implicated in a number of disease states, and topoisomerase poisons have proven to be effective anti-tumor drugs for some human malignancies. Topo I is mislocalized in Fanconi's anemia, and may be involved in the chromosomal breakage seen in this disorder (Wunder, E. (1984) Hum. Genet. 68:276-281). Overexpression of a truncated topo III in ataxia-telangiectasia (A-T) cells partially suppresses the A-T phenotype, probably through a dominant negative mechanism. This suggests that topo III is deregulated in A-T (Fritz, E. et al. (1997) Proc. Natl. Acad. Sci. USA 94:4538-4542). Topo III also interacts with the Bloom's Syndrome gene product, and has been suggested to have a role as a tumor suppressor (Wu, L. et al. (2000) J. Biol. Chem. 275:9636-9644). Aberrant topo II activity is often associated with cancer or increased cancer risk. Greatly lowered topo II activity has been found in some, but not all A-T cell lines (Mohamed, R. et al. (1987) Biochem. Biophys. Res. Commun. 149:233-238). On the other hand, topo II can break DNA in the region of the A-T gene (ATM), which controls all DNA damage-responsive cell cycle checkpoints (Kaufmann, W.K. (1998) Proc. Soc. Exp. Biol. Med. 217:327-334). The ability of topoisomerases to break DNA has been used as the basis of antitumor drugs. Topoisomerase poisons act by increasing the number of dead-end covalent DNA-enzyme complexes in the cell, ultimately triggering cell death pathways (Fortune, J.M. and N. Osheroff (2000) Prog. Nucleic Acid Res. Mol. Biol. 64:221-253; Guichard, S.M. and M.K. Danks (1999) Curr. Opin. Oncol. 11:482-489). Antibodies against topo I are found in the serum of systemic sclerosis patients, and the levels of the antibody may be used as a marker of pulmonary involvement in the disease (Diot, E. et al. (1999) Chest 116:715-720). Finally, the DNA binding region of human topo I has been used as a DNA delivery vehicle for gene therapy (Chen, T.Y. et al. (2000) Appl. Microbiol. Biotechnol. 53:558-567).

Recombinases

Genetic recombination is the process of rearranging DNA sequences within an organism's genome to provide genetic variation for the organism in response to changes in the environment. DNA recombination allows variation in the particular combination of genes present in an individual's genome, as well as the timing and level of expression of these genes. (See Alberts, *supra* pp. 263-273.) Two broad classes of genetic recombination are commonly recognized, general recombination and site-specific recombination. General recombination involves genetic exchange between any homologous pair of DNA sequences usually located on two copies of the same chromosome. The process is aided by enzymes, recombinases, that "nick" one strand of a DNA duplex more or less randomly and permit exchange with a complementary strand on another duplex. The process does not normally change the arrangement of genes in a chromosome. In site-specific recombination, the recombinase recognizes specific nucleotide sequences present in one or both of the recombining molecules. Base-pairing is not involved in this form of recombination and therefore it does not require DNA homology between the recombining molecules. Unlike general recombination, this form of recombination can alter the relative positions of nucleotide sequences in chromosomes.

RNA METABOLISM

Ribonucleic acid (RNA) is a linear single-stranded polymer of four nucleotides, ATP, CTP, UTP, and GTP. In most organisms, RNA is transcribed as a copy of deoxyribonucleic acid (DNA), the genetic material of the organism. In retroviruses RNA rather than DNA serves as the genetic material. RNA copies of the genetic material encode proteins or serve various structural, catalytic, or regulatory roles in organisms. RNA is classified according to its cellular localization and function. Messenger RNAs (mRNAs) encode polypeptides. Ribosomal RNAs (rRNAs) are assembled, along with ribosomal proteins, into ribosomes, which are cytoplasmic particles that translate mRNA into polypeptides. Transfer RNAs (tRNAs) are cytosolic adaptor molecules that function in mRNA translation by recognizing both an mRNA codon and the amino acid that matches that codon. Heterogeneous nuclear RNAs (hnRNAs) include mRNA precursors and other nuclear RNAs of various sizes. Small nuclear RNAs (snRNAs) are a part of the nuclear spliceosome complex that removes intervening, non-coding sequences (introns) and rejoins exons in pre-mRNAs.

Proteins are associated with RNA during its transcription from DNA, RNA processing, and translation of mRNA into protein. Proteins are also associated with RNA as it is used for structural, catalytic, and regulatory purposes.

RNA Processing

Ribosomal RNAs (rRNAs) are assembled, along with ribosomal proteins, into ribosomes, which are cytoplasmic particles that translate messenger RNA (mRNA) into polypeptides. The eukaryotic ribosome is composed of a 60S (large) subunit and a 40S (small) subunit, which together form the 80S ribosome. In addition to the 18S, 28S, 5S, and 5.8S rRNAs, ribosomes contain from 50

to over 80 different ribosomal proteins, depending on the organism. Ribosomal proteins are classified according to which subunit they belong (i.e., L, if associated with the large 60S large subunit or S if associated with the small 40S subunit). *E. coli* ribosomes have been the most thoroughly studied and contain 50 proteins, many of which are conserved in all life forms. The structures of nine ribosomal proteins have been solved to less than 3.0D resolution (i.e., S5, S6, S17, L1, L6, L9, L12, L14, L30), revealing common motifs, such as b-a-b protein folds in addition to acidic and basic RNA-binding motifs positioned between b-strands. Most ribosomal proteins are believed to contact rRNA directly (reviewed in Liljas, A. and M. Garber (1995) Curr. Opin. Struct. Biol. 5:721–727; see also Woodson, S.A. and N.B. Leontis (1998) Curr. Opin. Struct. Biol. 8:294-300; Ramakrishnan, V. and S.W. White (1998) Trends Biochem. Sci. 23:208-212).

Ribosomal proteins may undergo post-translational modifications or interact with other ribosome-associated proteins to regulate translation. For example, the highly homologous 40S ribosomal protein S6 kinases (S6K1 and S6K2) play a key role in the regulation of cell growth by controlling the biosynthesis of translational components which make up the protein synthetic apparatus (including the ribosomal proteins). In the case of S6K1, at least eight phosphorylation sites are believed to mediate kinase activation in a hierarchical fashion (Dufner and Thomas (1999) Exp. Cell. Res. 253:100-109). Some of the ribosomal proteins, including L1, also function as translational repressors by binding to polycistronic mRNAs encoding ribosomal proteins (reviewed in Liljas and Garber, *supra*).

Recent evidence suggests that a number of ribosomal proteins have secondary functions independent of their involvement in protein biosynthesis. These proteins function as regulators of cell proliferation and, in some instances, as inducers of cell death. For example, the expression of human ribosomal protein L13a has been shown to induce apoptosis by arresting cell growth in the G2/M phase of the cell cycle. Inhibition of expression of L13a induces apoptosis in target cells, which suggests that this protein is necessary, in the appropriate amount, for cell survival. Similar results have been obtained in yeast where inactivation of yeast homologues of L13a, rp22 and rp23, results in severe growth retardation and death. A closely related ribosomal protein, L7, arrests cells in G1 and also induces apoptosis. Thus, a subset of ribosomal proteins may function as cell cycle checkpoints and compose a new family of cell proliferation regulators.

Mapping of individual ribosomal proteins on the surface of intact ribosomes is accomplished using 3D immunocryoelectronmicroscopy, whereby antibodies raised against specific ribosomal proteins are visualized. Progress has been made toward the mapping of L1, L7, and L12 while the structure of the intact ribosome has been solved to only 20-25D resolution and inconsistencies exist among different crude structures (Frank, J. (1997) Curr. Opin. Struct. Biol. 7:266-272).

Three distinct sites have been identified on the ribosome. The aminoacyl-tRNA acceptor site

(A site) receives charged tRNAs (with the exception of the initiator-tRNA). The peptidyl-tRNA site (P site) binds the nascent polypeptide as the amino acid from the A site is added to the elongating chain. Deacylated tRNAs bind in the exit site (E site) prior to their release from the ribosome. The structure of the ribosome is reviewed in Stryer, L. (1995) Biochemistry, W.H. Freeman and Company, New York NY, pp. 888-9081; Lodish, *supra*, pp. 119-138; and Lewin, B (1997) Genes VI, Oxford University Press, Inc. New York, NY).

Various proteins are necessary for processing of transcribed RNAs in the nucleus. Pre-mRNA processing steps include capping at the 5' end with methylguanosine, polyadenylating the 3' end, and splicing to remove introns. The primary RNA transript from DNA is a faithful copy of the gene containing both exon and intron sequences, and the latter sequences must be cut out of the RNA transcript to produce a mRNA that codes for a protein. This "splicing" of the mRNA sequence takes place in the nucleus with the aid of a large, multicomponent ribonucleoprotein complex known as a spliceosome. The spliceosomal complex is comprised of five small nuclear ribonucleoprotein particles (snRNPs) designated U1, U2, U4, U5, and U6. Each snRNP contains a single species of snRNA and about ten proteins. The RNA components of some snRNPs recognize and base-pair with intron consensus sequences. The protein components mediate spliceosome assembly and the splicing reaction. Autoantibodies to snRNP proteins are found in the blood of patients with systemic lupus erythematosus (Stryer, *supra*, p. 863).

Heterogeneous nuclear ribonucleoproteins (hnRNPs) have been identified that have roles in splicing, exporting of the mature RNAs to the cytoplasm, and mRNA translation (Biamonti, G. et al. (1998) Clin. Exp. Rheumatol. 16:317-326). Some examples of hnRNPs include the yeast proteins Hrp1p, involved in cleavage and polyadenylation at the 3' end of the RNA; Cbp80p, involved in capping the 5' end of the RNA; and Npl3p, a homolog of mammalian hnRNP A1, involved in export of mRNA from the nucleus (Shen, E.C. et al. (1998) Genes Dev. 12:679-691). HnRNPs have been shown to be important targets of the autoimmune response in rheumatic diseases (Biamonti, *supra*).

Many snRNP and hnRNP proteins are characterized by an RNA recognition motif (RRM). (Reviewed in Birney, E. et al. (1993) Nucleic Acids Res. 21:5803-5816.) The RRM is about 80 amino acids in length and forms four b-strands and two a-helices arranged in an a /b sandwich. The RRM contains a core RNP-1 octapeptide motif along with surrounding conserved sequences. In addition to snRNP proteins, examples of RNA-binding proteins which contain the above motifs include heteronuclear ribonucleoproteins which stabilize nascent RNA and factors which regulate alternative splicing. Alternative splicing factors include developmentally regulated proteins, specific examples of which have been identified in lower eukaryotes such as *Drosophila melanogaster* and *Caenorhabditis elegans*. These proteins play key roles in developmental processes such as pattern formation and sex determination, respectively. (See, for example, Hodgkin, J. et al. (1994)

Development 120:3681-3689.)

The 3' ends of most eukaryote mRNAs are also posttranscriptionally modified by polyadenylation. Polyadenylation proceeds through two enzymatically distinct steps: (i) the endonucleolytic cleavage of nascent mRNAs at *cis*-acting polyadenylation signals in the 3'-untranslated (non-coding) region and (ii) the addition of a poly(A) tract to the 5' mRNA fragment. The presence of *cis*-acting RNA sequences is necessary for both steps. These sequences include 5'-AAUAAA-3' located 10-30 nucleotides upstream of the cleavage site and a less well-conserved GU- or U-rich sequence element located 10-30 nucleotides downstream of the cleavage site. Cleavage stimulation factor (CstF), cleavage factor I (CF I), and cleavage factor II (CF II) are involved in the cleavage reaction while cleavage and polyadenylation specificity factor (CPSF) and poly(A) polymerase (PAP) are necessary for both cleavage and polyadenylation. An additional enzyme, poly(A)-binding protein II (PAB II), promotes poly(A) tract elongation (Rüegsegger, U. et al. (1996) J. Biol. Chem. 271:6107-6113; and references within).

YT521-B is a nuclear protein that was identified by using a yeast two-hybrid screen for proteins that interact with known mRNA splicing factors (Hartmann, A.M. et al. (1999) Mol. Biol. Cell 10:3909-3926). The protein contains four nuclear localization signals, an N-terminal glutamic acid-rich region, a glutamic acid/arginine-rich region, and a C-terminal proline-rich region. YT521 associates with the nuclear transcriptosomal component scaffold attachment factor B and with the Src kinase substrate, Sam68. Phosphorylation of Sam68 by Src family kinase $p59^{fyn}$ reduces the association of Sam68 with YT521-B. Both YT521 and Sam68 may participate in a signal transduction pathway that controls alternative splice site selection.

TRANSLATION

Correct translation of the genetic code depends upon each amino acid forming a linkage with the appropriate transfer RNA (tRNA). The aminoacyl-tRNA synthetases (aaRSs) are essential proteins found in all living organisms. The aaRSs are responsible for the activation and correct attachment of an amino acid with its cognate tRNA, as the first step in protein biosynthesis. Prokaryotic organisms have at least twenty different types of aaRSs, one for each different amino acid, while eukaryotes usually have two aaRSs, a cytosolic form and a mitochondrial form, for each different amino acid. The 20 aaRS enzymes can be divided into two structural classes. Class I enzymes add amino acids to the 2' hydroxyl at the 3' end of tRNAs while Class II enzymes add amino acids to the 3' hydroxyl at the 3' end of tRNAs. Each class is characterized by a distinctive topology of the catalytic domain. Class I enzymes contain a catalytic domain based on the nucleotide-binding Rossman 'fold'. In particular, a consensus tetrapeptide motif is highly conserved (Prosite Document PDOC00161, Aminoacyl-transfer RNA synthetases class-I signature). Class I enzymes are specific for arginine, cysteine, glutamic acid, glutamine, isoleucine, leucine, methionine, tyrosine, tryptophan,

and valine. Class II enzymes contain a central catalytic domain, which consists of a seven-stranded antiparallel ß-sheet domain, as well as N- and C- terminal regulatory domains. Class II enzymes are separated into two groups based on the heterodimeric or homodimeric structure of the enzyme; the latter group is further subdivided by the structure of the N- and C-terminal regulatory domains (Hartlein, M. and S. Cusack (1995) J. Mol. Evol. 40:519-530). Class II enzymes are specific for alanine, asparagine, aspartic acid, glycine, histidine, lysine, phenylalanine, proline, serine, and threonine.

Certain aaRSs also have editing functions. IleRS, for example, can misactivate valine to form Val-tRNA^{Ile}, but this product is cleared by a hydrolytic activity that destroys the mischarged product. This editing activity is located within a second catalytic site found in the connective polypeptide 1 region (CP1), a long insertion sequence within the Rossman fold domain of Class I enzymes (Schimmel, P. et al. (1998) FASEB J. 12:1599-1609). AaRSs also play a role in tRNA processing. It has been shown that mature tRNAs are charged with their respective amino acids in the nucleus before export to the cytoplasm, and charging may serve as a quality control mechanism to insure the tRNAs are functional (Martinis, S.A. et al. (1999) EMBO J. 18:4591-4596).

Under optimal conditions, polypeptide synthesis proceeds at a rate of approximately 40 amino acid residues per second. The rate of misincorporation during translation in on the order of 10^{-4} and is primarily the result of aminoacyl-t-RNAs being charged with the incorrect amino acid. Incorrectly charged tRNA are toxic to cells as they result in the incorporation of incorrect amino acid residues into an elongating polypeptide. The rate of translation is presumed to be a compromise between the optimal rate of elongation and the need for translational fidelity. Mathematical calculations predict that 10^{-4} is indeed the maximum acceptable error rate for protein synthesis in a biological system (reviewed in Stryer, *supra*; and Watson, J. et al. (1987) The Benjamin/Cummings Publishing Co., Inc. Menlo Park, CA). A particularly error prone aminoacyl-tRNA charging event is the charging of tRNA^{Gln} with Gln. A mechanism exits for the correction of this mischarging event which likely has its origins in evolution. Gln was among the last of the 20 naturally occurring amino acids used in polypeptide synthesis to appear in nature. Gram positive eubacteria, cyanobacteria, Archeae, and eukaryotic organelles possess a noncanonical pathway for the synthesis of Gln-tRNA^{Gln} based on the transformation of Glu-tRNA^{Gln} (synthesized by Glu-tRNA synthetase, GluRS) using the enzyme Glu-tRNA^{Gln} amidotransferase (Glu-AdT). The reactions involved in the transamidation pathway are as follows (Curnow, A.W. et al. (1997) Nucleic Acids Symposium 36:2-4):

$$tRNA^{Gln} + Glu + ATP \xrightarrow{GluRS} Glu\text{-}tRNA^{Gln} + AMP + PP_i$$

Glu-AdT

$$Glu\text{-}tRNA^{Gln} + Gln + ATP \Rightarrow Gln\text{-}tRNA^{Gln} + Glu + ADP + P$$

A similar enzyme, Asp-$tRNA^{Asn}$ amidotransferase, exists in Archaea, which transforms Asp-$tRNA^{Asn}$ to Asn-$tRNA^{Asn}$. Formylase, the enzyme that transforms Met-$tRNA^{fMet}$ to fMet-$tRNA^{fMet}$ in eubacteria, is likely to be a related enzyme. A hydrolytic activity has also been identified that destroys mischarged Val-$tRNA^{Ile}$ (Schimmel, P. et al. (1998) FASEB J. 12:1599-1609). One likely scenario for the evolution of Glu-AdT in primitive life forms is the absence of a specific glutaminyl-tRNA synthetase (GlnRS), requiring an alternative pathway for the synthesis of Gln-$tRNA^{Gln}$. In fact, deletion of the Glu-AdT operon in Gram positive bacteria is lethal (Curnow, A.W. et al. (1997) Proc. Natl. Acad. Sci. USA 94:11819-11826). The existence of GluRS activity in other organisms has been inferred by the high degree of conservation in translation machinery in nature; however, GluRS has not been identified in all organisms, including *Homo sapiens*. Such an enzyme would be responsible for ensuring translational fidelity and reducing the synthesis of defective polypeptides.

The different aaRSs are believed to be the result of divergent evolution, likely following gene duplication events. Notably, amino acids such as Gln, were among the last to appear in nature and evolutionary studies suggest that Gln-RSs appeared first in eukaryotes and were later horizontally transferred to prokaryotes (Lamour, V. et al. (1994) Proc. Natl. Acad. Sci. U.S.A. 91:8670-74 and Siatecka, M. et al. (1998) Eur. J. Biochem. 256:80-7). The importance of Gln-RS and Gln-$tRNA^{Gln}$ are discussed below.

In addition to their function in protein synthesis, specific aminoacyl tRNA synthetases also play roles in cellular fidelity, RNA splicing, RNA trafficking, apoptosis, and transcriptional and translational regulation. For example, human tyrosyl-tRNA synthetase can be proteolytically cleaved into two fragments with distinct cytokine activities. The carboxy-terminal domain exhibits monocyte and leukocyte chemotaxis activity as well as stimulating production of myeloperoxidase, tumor necrosis factor-a, and tissue factor. The N-terminal domain binds to the interleukin-8 type A receptor and functions as an interleukin-8-like cytokine. Human tyrosyl-tRNA synthetase is secreted from apoptotic tumor cells and may accelerate apoptosis (Wakasugi, K., and Schimmel, P. (1999) Science 284:147-151). Mitochondrial *Neurospora crassa* TyrRS and *S. cerevisiae* LeuRS are essential factors for certain group I intron splicing activities, and human mitochondrial LeuRS can substitute for the yeast LeuRS in a yeast null strain. Certain bacterial aaRSs are involved in regulating their own transcription or translation (Martinis, *supra*). Several aaRSs are able to synthesize diadenosine oligophosphates, a class of signalling molecules with roles in cell proliferation, differentiation, and apoptosis (Kisselev, L.L et al. (1998) FEBS Lett. 427:157-163; Vartanian, A. et al. (1999) FEBS Lett. 456:175-180).

Autoantibodies against aminoacyl-tRNAs are generated by patients with autoimmune

diseases such as rheumatic arthritis, dermatomyositis and polymyositis, and correlate strongly with complicating interstitial lung disease (ILD) (Freist, W. et al. (1999) Biol. Chem. 380:623-646; Freist, W. et al. (1996) Biol. Chem. Hoppe Seyler 377:343-356). These antibodies appear to be generated in response to viral infection, and coxsackie virus has been used to induce experimental viral myositis in animals.

Comparison of aaRS structures between humans and pathogens has been useful in the design of novel antibiotics (Schimmel, *supra*). Genetically engineered aaRSs have been utilized to allow site-specific incorporation of unnatural amino acids into proteins *in vivo* (Liu, D.R. et al. (1997) Proc. Natl. Acad. Sci. USA 94:10092-10097).

tRNA Modifications

The modified ribonucleoside, pseudouridine (y), is present ubiquitously in the anticodon regions of transfer RNAs (tRNAs), large and small ribosomal RNAs (rRNAs), and small nuclear RNAs (snRNAs). y is the most common of the modified nucleosides (i.e., other than G, A, U, and C) present in tRNAs. Only a few yeast tRNAs that are not involved in protein synthesis do not contain y (Cortese, R. et al. (1974) J. Biol. Chem. 249:1103-1108). The enzyme responsible for the conversion of uridine to y, pseudouridine synthase (pseudouridylate synthase), was first isolated from *Salmonella typhimurium* (Arena, F. et al. (1978) Nucleic Acids Res. 5:4523-4536). The enzyme has since been isolated from a number of mammals, including steer and mice (Green, C.J. et al. (1982) J. Biol. Chem. 257:3045-52; and Chen, J. and J.R. Patton (1999) RNA 5:409-419). tRNA pseudouridine synthases have been the most extensively studied members of the family. They require a thiol donor (e.g., cysteine) and a monovalent cation (e.g., ammonia or potassium) for optimal activity. Additional cofactors or high energy molecules (e.g., ATP or GTP) are not required (Green et al., *supra*). Other eukaryotic pseudouridine synthases have been identified that appear to be specific for rRNA (reviewed in Smith, C.M. and J.A. Steitz (1997) Cell 89:669-672) and a dual-specificity enzyme has been identified that uses both tRNA and rRNA substrates (Wrzesinski, J. et al. (1995) RNA 1: 437-448). The absence of y in the anticodon loop of tRNAs results in reduced growth in both bacteria (Singer, C.E. et al. (1972) Nature New Biol. 238:72-74) and yeast (Lecointe, F. (1998) J. Biol. Chem. 273:1316-1323), although the genetic defect is not lethal.

Another ribonucleoside modification that occurs primarily in eukaryotic cells is the conversion of guanosine to N^2,N^2-dimethylguanosine (m^2_2G) at position 26 or 10 at the base of the D-stem of cytosolic and mitochondrial tRNAs. This posttranscriptional modification is believed to stabilize tRNA structure by preventing the formation of alternative tRNA secondary and tertiary structures. Yeast $tRNA^{Asp}$ is unusual in that it does not contain this modification. The modification does not occur in eubacteria, presumably because the structure of tRNAs in these cells and organelles is sequence constrained and does not require posttranscriptional modification to prevent the formation

of alternative structures (Steinberg, S. and R. Cedergren (1995) RNA 1:886-891, and references within). The enzyme responsible for the conversion of guanosine to m^2_2G is a 63 kDa *S*-adenosylmethionine (SAM)-dependent tRNA N^2,N^2-dimethyl-guanosine methyltransferase (also referred to as the *TRM1* gene product and herein referred to as TRM) (Edqvist, J. (1995) Biochimie 77:54-61). The enzyme localizes to both the nucleus and the mitochondria (Li, J-M. et al. (1989) J. Cell Biol. 109:1411-1419). Based on studies with TRM from *Xenopus laevis*, there appears to be a requirement for base pairing at positions C11-G24 and G10-C25 immediately preceding the G26 to be modified, with other structural features of the tRNA also being required for the proper presentation of the G26 substrate (Edqvist. J. et al. (1992) Nucleic Acids Res. 20:6575-6581). Studies in yeast suggest that cells carrying a weak ochre tRNA suppressor (sup3-i) are unable to suppress translation termination in the absence of TRM activity, suggesting a role for TRM in modifying the frequency of suppression in eukaryotic cells (Niederberger, C. et al. (1999) FEBS Lett. 464:67-70), in addition to the more general function of ensuring the proper three-dimensional structures for tRNA.

Translation Initiation

Initiation of translation can be divided into three stages. The first stage brings an initiator transfer RNA (Met-$tRNA_f$) together with the 40S ribosomal subunit to form the 43S preinitiation complex. The second stage binds the 43S preinitiation complex to the mRNA, followed by migration of the complex to the correct AUG initiation codon. The third stage brings the 60S ribosomal subunit to the 40S subunit to generate an 80S ribosome at the inititation codon. Regulation of translation primarily involves the first and second stage in the initiation process (Pain, V.M. (1996) Eur. J. Biochem. 236:747-771).

Several initiation factors, many of which contain multiple subunits, are involved in bringing an initiator tRNA and the 40S ribosomal subunit together. eIF2, a guanine nucleotide binding protein, recruits the initiator tRNA to the 40S ribosomal subunit. Only when eIF2 is bound to GTP does it associate with the initiator tRNA. eIF2B, a guanine nucleotide exchange protein, is responsible for converting eIF2 from the GDP-bound inactive form to the GTP-bound active form. Two other factors, eIF1A and eIF3 bind and stabilize the 40S subunit by interacting with the 18S ribosomal RNA and specific ribosomal structural proteins. eIF3 is also involved in association of the 40S ribosomal subunit with mRNA. The Met-$tRNA_f$, eIF1A, eIF3, and 40S ribosomal subunit together make up the 43S preinitiation complex (Pain, *supra*).

Additional factors are required for binding of the 43S preinitiation complex to an mRNA molecule, and the process is regulated at several levels. eIF4F is a complex consisting of three proteins: eIF4E, eIF4A, and eIF4G. eIF4E recognizes and binds to the mRNA 5'-terminal m^7GTP cap, eIF4A is a bidirectional RNA-dependent helicase, and eIF4G is a scaffolding polypeptide. eIF4G has three binding domains. The N-terminal third of eIF4G interacts with eIF4E, the central

third interacts with eIF4A, and the C-terminal third interacts with eIF3 bound to the 43S preinitiation complex. Thus, eIF4G acts as a bridge between the 40S ribosomal subunit and the mRNA (Hentze, M.W. (1997) Science 275:500-501).

The ability of eIF4F to initiate binding of the 43S preinitiation complex is regulated by structural features of the mRNA. The mRNA molecule has an untranslated region (UTR) between the 5' cap and the AUG start codon. In some mRNAs this region forms secondary structures that impede binding of the 43S preinitiation complex. The helicase activity of eIF4A is thought to function in removing this secondary structure to facilitate binding of the 43S preinitiation complex (Pain, *supra*).

The translation of eukaryotic mRNA is a highly competitive and tightly regulated step in gene expression. Control of this step is most commonly exerted at the rate-limiting initiation phase. Ribosomal proteins involved in translation initiation have been known for some time and their biochemical activities were used to build the currently accepted model for cap-dependent initiation of translation (Merrick, W. C. et al. (1996) in *Translational Control*, Hershey, J. W. B. et al. *Ed.*, Cold Spring Harbor Laboratory Press, pp. 31-69). According to this model, the 5′ cap structure (m^7GpppN) attracts the eukaryotic initiation factor 4F (eIF4F) complex to the mRNA. eIF4F is a heteromultimeric complex composed of the cap-binding protein eIF4E, the RNA-dependent ATPase eIF4A, and the modular factor eIF4G. The small (40S) ribosomal subunit binds to the 5′ end of an mRNA as a 43S complex which is thought to unwind secondary structure in the 5′ UTR. The resulting 48S complex then advances through the initiation cycle. A later movement of the 43S complex along the mRNA, termed scanning, is the most plausible explanation for a faithful recognition of the (usually) first AUG triplet as the start codon. Codon-anticodon base-pairing with Met-$tRNA^i$ triggers eukaryotic initiation factor 2 (eIF2)-bound GTP hydrolysis, catalysed by eukaryotic initiation factor 5 (eIF5). It has been thought that this causes dissociation of initiation factors and the large (60S) subunit joining to form the 80S ribosome.

The bacterial translation initiation factor, IF2, is found to be evolutionarily conserved with homologs identified in archae, yeasts, mammals, zebrafish, and maize (Choi, S. D. et al. (1998) Science 280:1757-1760; Lee, J. H. et al. (1999) Proc. Natl. Acad. Sci. U.S.A 96:4342-4347). Mutant strains of Saccharomyces cerevisiae which lack the gene which encodes yeast IF2 can be used to demonstrate this evolutionary conservation with respect to IF2 activity. Protein biosynthetic activity of translation extracts prepared from such mutant strains can be restored by addition of recombinant yIF2 as described in Choi et al. (supra). Evidence that the biologic activity of these same translation extracts can be restored by addition of either human or archeal IF2 (Lee et al. supra), supports the idea of universal conservation of IF2 function throughout evolution.

The eukaryotic translation initiation factor 4E (eIF4E) regulates the rate of translation

initiation. Overexpression of eIF4E results in rapid cell or tissue proliferation and malignant transformation. eIF4E facilitates the synthesis of two powerful tumor angiogenic factors (VEGF and FGF-2) by selectively enhancing their translation. eIF4E is overexpressed not only in all head and neck squamous cell cancers but also in some dysplastic margins. Tumorigenesis in the head and neck is proposed to be a multistep process preceded by clinically evident precancerous lesions (Nathan, C.-A. O. et al. (1999) Laryngoscope 109:1253-1258; De Benedetti, A. and A. L. Harris (1999) Int. J. Biochem. Cell Biol. 31:59-72).

The human eukaryotic protein translation initiation factor, eIF2, binds GTP and Met-tRNAi then transfers Met-tRNAi to the 40S ribosomal subunit in a rate-limiting step in mRNA translation. One member of this highly conserved, multigene family is the human eIF2C1 gene. This gene has been mapped to chromosome 1p34-p35, which is a genomic area often lost in human cancers such as Wilms tumors, neuroblastoma, and carcinomas of the breast, liver, and colon (Koesters, R. (1999) Genomics 61:210-218).

Elongation factor 2 (eEF-2) is a 100-kDa protein that catalyzes the ribosomal translocation reaction, resulting in the movement of ribosomes along mRNA. eEF-2 is the target for a very specific Ca^{2+}/calmodulin-dependent eEF-2 kinase. Phosphorylation of eEF-2 makes it inactive in translation, which suggests that protein synthesis can be regulated by Ca^{2+} through eEF-2 phosphorylation. eEF-2 phosphorylation therefore regulates the cell-cycle and other processes where changes of intracellular Ca^{2+} concentration induce a new physiological state of a cell. The main role of eEF-2 phosphorylation in these processes is temporary inhibition of overall translation in response to transient elevation of the Ca^{2+} concentrations in the cytoplasm. Temporary inhibition of translation may trigger the transition of a cell from one physiologic state into another because of the disappearance of short-lived repressors and thus the activation of expression of new genes (Ryazanov, A. G. and A. S. Spirin (1990) New Biol 2:843-850).

Other ribosomal proteins which modulate translation of mRNA include the retinoblastoma protein (Rb1), HIV-1 TAR RNA binding protein (TARBP-b), v-fos transformation effector protein (Fte-1), the colin carcinoma laminin-binding protein, the Wilm's tumor-related protein (QM), the ribosomal phosphoproteins P0, P1, and P2, ubiquitin, and the Epstein-Barr virus small RNAs-associated protein (EAP).

Translation Elongation

Elongation is the process whereby additional amino acids are joined to the initiator methionine to form the complete polypeptide chain. The elongation factors EF1 a, EF1 b g, and EF2 are involved in elongating the polypeptide chain following initiation. EF1 a is a GTP-binding protein. In EF1 a's GTP-bound form, it brings an aminoacyl-tRNA to the ribosome's A site. The amino acid attached to the newly arrived aminoacyl-tRNA forms a peptide bond with the initiatior methionine.

The GTP on EF1 a is hydrolyzed to GDP, and EF1 a -GDP dissociates from the ribosome. EF1 b g binds EF1 a -GDP and induces the dissociation of GDP from EF1 a, allowing EF1 a to bind GTP and a new cycle to begin.

As subsequent aminoacyl-tRNAs are brought to the ribosome, EF-G, another GTP-binding protein, catalyzes the translocation of tRNAs from the A site to the P site and finally to the E site of the ribosome. This allows the ribosome and the mRNA to remain attached during translation.

Translation Termination

The release factor eRF carries out termination of translation. eRF recognizes stop codons in the mRNA, leading to the release of the polypeptide chain from the ribosome.

Expression profiling

Microarrays are analytical tools used in bioanalysis. A microarray has a plurality of molecules spatially distributed over, and stably associated with, the surface of a solid support. Microarrays of polypeptides, polynucleotides, and/or antibodies have been developed and find use in a variety of applications, such as gene sequencing, monitoring gene expression, gene mapping, bacterial identification, drug discovery, and combinatorial chemistry.

One area in particular in which microarrays find use is in gene expression analysis. Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder.

There is a need in the art for new compositions, including nucleic acids and proteins, for the diagnosis, prevention, and treatment of cell proliferative, neurological, developmental, and autoimmune/inflammatory disorders, and infections.

SUMMARY OF THE INVENTION

Various embodiments of the invention provide purified polypeptides, nucleic acid-associated proteins, referred to collectively as "NAAP" and individually as "NAAP-1," "NAAP-2," "NAAP-3," "NAAP-4," "NAAP-5," "NAAP-6," "NAAP-7," "NAAP-8," "NAAP-9," "NAAP-10," "NAAP-11," "NAAP-12," "NAAP-13," "NAAP-14," "NAAP-15," "NAAP-16," "NAAP-17, "NAAP-18," "NAAP-19," "NAAP-20," "NAAP-21," "NAAP-22," "NAAP-23," "NAAP-24," "NAAP-25," "NAAP-26," "NAAP-27," "NAAP-28," "NAAP-29," "NAAP-30," "NAAP-31," "NAAP-32," "NAAP-33," "NAAP-34," and "NAAP-35," and methods for using these proteins and their encoding

polynucleotides for the detection, diagnosis, and treatment of diseases and medical conditions. Embodiments also provide methods for utilizing the purified nucleic acid-associated proteins and/or their encoding polynucleotides for facilitating the drug discovery process, including determination of efficacy, dosage, toxicity, and pharmacology. Related embodiments provide methods for utilizing the purified nucleic acid-associated proteins and/or their encoding polynucleotides for investigating the pathogenesis of diseases and medical conditions.

An embodiment provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-35. Another embodiment provides an isolated polypeptide comprising an amino acid sequence of SEQ ID NO:1-35.

Still another embodiment provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-35. In another embodiment, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-35. In an alternative embodiment, the polynucleotide is selected from the group consisting of SEQ ID NO:36-70.

Still another embodiment provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-35. Another embodiment provides a cell transformed with the recombinant polynucleotide. Yet another embodiment provides a transgenic organism comprising the recombinant polynucleotide.

Another embodiment provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-35. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Yet another embodiment provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-35.

Still yet another embodiment provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:36-70, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:36-70, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In other embodiments, the polynucleotide can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:36-70, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:36-70, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides

comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex. In a related embodiment, the method can include detecting the amount of the hybridization complex. In still other embodiments, the probe can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Still yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:36-70, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:36-70, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof. In a related embodiment, the method can include detecting the amount of the amplified target polynucleotide or fragment thereof.

Another embodiment provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, and a pharmaceutically acceptable excipient. In one embodiment, the composition can comprise an amino acid sequence selected from the group consisting of SEQ ID NO:1-35. Other embodiments provide a method of treating a disease or condition associated with decreased or abnormal expression of functional NAAP, comprising administering to a patient in need of such treatment the composition.

Yet another embodiment provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ

ID NO:1-35, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-35. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. Another embodiment provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with decreased expression of functional NAAP, comprising administering to a patient in need of such treatment the composition.

Still yet another embodiment provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-35. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. Another embodiment provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with overexpression of functional NAAP, comprising administering to a patient in need of such treatment the composition.

Another embodiment provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-35. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

Yet another embodiment provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an

amino acid sequence selected from the group consisting of SEQ ID NO:1-35, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-35. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

Still yet another embodiment provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:36-70, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

Another embodiment provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:36-70, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:36-70, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:36-70, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:36-70, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide can comprise a fragment of a polynucleotide selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d)

comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for full length polynucleotide and polypeptide embodiments of the invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptide embodiments of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide embodiments, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide embodiments, along with selected fragments of the polynucleotides.

Table 5 shows representative cDNA libraries for polynucleotide embodiments.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze polynucleotides and polypeptides, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleic acids, and methods are described, it is understood that embodiments of the invention are not limited to the particular machines, instruments, materials, and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the invention.

As used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now

described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with various embodiments of the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"NAAP" refers to the amino acid sequences of substantially purified NAAP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of NAAP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of NAAP either by directly interacting with NAAP or by acting on components of the biological pathway in which NAAP participates.

An "allelic variant" is an alternative form of the gene encoding NAAP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding NAAP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as NAAP or a polypeptide with at least one functional characteristic of NAAP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding NAAP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide encoding NAAP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent NAAP. Deliberate amino acid substitutions may be made on the basis of one or more similarities in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of NAAP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and

valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" can refer to an oligopeptide, a peptide, a polypeptide, or a protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid. Amplification may be carried out using polymerase chain reaction (PCR) technologies or other nucleic acid amplification technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of NAAP. Antagonists may include proteins such as antibodies, anticalins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of NAAP either by directly interacting with NAAP or by acting on components of the biological pathway in which NAAP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind NAAP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include

deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH_2), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker (Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13).

The term "intramer" refers to an aptamer which is expressed *in vivo*. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl. Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a polynucleotide having a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic NAAP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide" and a "composition comprising a given polypeptide" can refer to any composition containing the given polynucleotide or polypeptide. The

composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotides encoding NAAP or fragments of NAAP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of

the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of NAAP or a polynucleotide encoding NAAP which can be identical in sequence to, but shorter in length than, the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from about 5 to about 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:36-70 can comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:36-70, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:36-70 can be employed in one or more embodiments of methods of the invention, for example, in hybridization and

amplification technologies and in analogous methods that distinguish SEQ ID NO:36-70 from related polynucleotides. The precise length of a fragment of SEQ ID NO:36-70 and the region of SEQ ID NO:36-70 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-35 is encoded by a fragment of SEQ ID NO:36-70. A fragment of SEQ ID NO:1-35 can comprise a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-35. For example, a fragment of SEQ ID NO:1-35 can be used as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-35. The precise length of a fragment of SEQ ID NO:1-35 and the region of SEQ ID NO:1-35 to which the fragment corresponds can be determined based on the intended purpose for the fragment using one or more analytical methods described herein or otherwise known in the art.

A "full length" polynucleotide is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using one or more computer algorithms or programs known in the art or described herein. For example, percent identity can be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989; CABIOS 5:151-153) and in Higgins, D.G. et al. (1992; CABIOS 8:189-191). For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms which can be used is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410),

which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the

stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acids by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid present in solution and another nucleic acid immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or polynucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune

disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of NAAP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of NAAP which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, antibodies, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, antibody, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of NAAP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of NAAP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an NAAP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of NAAP.

"Probe" refers to nucleic acids encoding NAAP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acids. Probes are isolated

oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989; Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY), Ausubel, F.M. et al. (1999) Short Protocols in Molecular Biology, 4th ed., John Wiley & Sons, New York NY), and Innis, M. et al. (1990; PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved

regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a nucleic acid that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, *supra*. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA molecule, is composed of the same linear sequence of nucleotides as the reference DNA molecule with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing NAAP, nucleic acids encoding NAAP, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a

protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably at least about 75% free, and most preferably at least about 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. In another embodiment, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather

is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), *supra*.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotides that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

Various embodiments of the invention include new human nucleic acid-associated proteins (NAAP), the polynucleotides encoding NAAP, and the use of these compositions for the diagnosis,

treatment, or prevention of cell proliferative, neurological, developmental, and autoimmune/inflammatory disorders, and infections.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide embodiments of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown. Column 6 shows the Incyte ID numbers of physical, full length clones corresponding to polypeptide and polynucleotide embodiments. The full length clones encode polypeptides which have at least 95% sequence identity to the polypeptides shown in column 3.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are nucleic acid-associated proteins. For example, SEQ ID NO:1 is 80% identical, from residue P24 to residue T316, to Rattus rattus ribosomal protein S2 (GenBank ID g57718) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 2.2e-116, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains a ribosomal protein

S5 domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and PROFILESCAN analyses and BLAST analyses of the PRODOM and DOMO databases provide further corroborative evidence that SEQ ID NO:1 is a ribosomal protein. In another example, SEQ ID NO:4 is 97% identical, from residue S650 to residue R1142, to human CAGH32 (GenBank ID g2565061) as determined by BLAST. The BLAST probability score is 1.8e-254. SEQ ID NO:4 also contains a helicase conserved C-terminal domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database . Data from MOTIFS analysis and BLAST analysis of the PRODOM and DOMO databases provide further corroborative evidence that SEQ ID NO:4 is a DNA modification enzyme such as a helicase. In another example, SEQ ID NO:10 is 91% identical from residue V59 to residue E290, and 58% identical from residue M1 to residue P275, to Bos taurus transcription factor EF1(A) (GenBank ID g162983) as determined BLAST. The BLAST probability score from residue V59 to residue E290 is 6.7e-115. SEQ ID NO:10 also contains a "cold shock" DNA-binding domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database. Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:10 is a transcription factor. In another example, SEQ ID NO:22 is 88% identical, from residue M1 to residue H451, to human testis-specific RING finger protein (GenBank ID g9650982) as determined by BLAST. The BLAST probability score is 1.8e-215. SEQ ID NO:22 also contains SPRY, B-box zinc-finger, and zinc-finger type C3HC4 domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database. Data from BLAST-DOMO and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:22 is a RFP transforming protein or cell attachment sequence. In another example, SEQ ID NO:25 is 95% identical, from residue M1 to residue D334, to mouse ventral anterior homeobox-containing protein-1 (GenBank ID g3641258) as determined by BLAST. The BLAST probability score is 2.7e-166. SEQ ID NO:25 also contains a homeobox domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database. Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:25 is a homeobox-containing protein. In a further example, SEQ ID NO:31 is 90% identical, from residue R130 to residue A816, and 68% identical, from residue M1 to residue S129 to human eukaryotic initiation factor, EIF2C1, (GenBank ID g6002623) as determined by BLAST. The BLAST probability score is 0.0. SEQ ID NO:31 also contains a PAZ (proteins Piwi, Argonaut, and Zwille/Pinhead) domain and a Piwi (a Drosophila protein which functions in RNA interference) domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM. In yet another example, SEQ ID NO:33 is 74% identical, from

residue M1 to residue Y255, to human zinc finger protein (GenBank ID g347906) as determined by BLAST. The BLAST probability score is 3.6e-100. SEQ ID NO:33 also contains a KRAB box as well as a zinc finger domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database. Data from MOTIFS and further BLAST analyses provide further corroborative evidence that SEQ ID NO:33 is a zinc finger protein. SEQ ID NO:2-3, SEQ ID NO:5-9, SEQ ID NO:11-21, SEQ ID NO:23-24, SEQ ID NO:26-30, SEQ ID NO:32, and SEQ ID NO:34-35 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-35 are described in Table 7.

As shown in Table 4, the full length polynucleotide embodiments were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5′) and stop (3′) positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide embodiments, and of fragments of the polynucleotides which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:36-70 or that distinguish between SEQ ID NO:36-70 and related polynucleotides.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotides. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL_*XXXXXX*_N_1_N_2_*YYYYY*_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an

"exon-stretching" algorithm. For example, a polynucleotide sequence identified as FL*XXXXXX_gAAAAA_gBBBBB_1_N* is a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g*BBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	**Type of analysis and/or examples of programs**
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotides which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotides. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses NAAP variants. A preferred NAAP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the NAAP amino acid sequence, and which contains at least one functional or structural characteristic of NAAP.

Various embodiments also encompass polynucleotides which encode NAAP. In a particular

embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:36-70, which encodes NAAP. The polynucleotide sequences of SEQ ID NO:36-70, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses variants of a polynucleotide encoding NAAP. In particular, such a variant polynucleotide will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a polynucleotide encoding NAAP. A particular aspect of the invention encompasses a variant of a polynucleotide comprising a sequence selected from the group consisting of SEQ ID NO:36-70 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:36-70. Any one of the polynucleotide variants described above can encode a polypeptide which contains at least one functional or structural characteristic of NAAP.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide encoding NAAP. A splice variant may have portions which have significant sequence identity to a polynucleotide encoding NAAP, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to a polynucleotide encoding NAAP over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide encoding NAAP. Any one of the splice variants described above can encode a polypeptide which contains at least one functional or structural characteristic of NAAP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding NAAP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring NAAP, and all such variations are to be considered as being specifically disclosed.

Although polynucleotides which encode NAAP and its variants are generally capable of hybridizing to polynucleotides encoding naturally occurring NAAP under appropriately selected

conditions of stringency, it may be advantageous to produce polynucleotides encoding NAAP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding NAAP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of polynucleotides which encode NAAP and NAAP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic polynucleotide may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a polynucleotide encoding NAAP or any fragment thereof.

Embodiments of the invention can also include polynucleotides that are capable of hybridizing to the claimed polynucleotides, and, in particular, to those having the sequences shown in SEQ ID NO:36-70 and fragments thereof, under various conditions of stringency (Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511). Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Biosciences, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Invitrogen, Carlsbad CA). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Amersham Biosciences), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art (Ausubel et al., *supra*, ch. 7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853).

The nucleic acids encoding NAAP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed,

restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186). A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art (Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotides or fragments thereof which encode NAAP may be cloned in recombinant DNA molecules that direct expression of NAAP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other polynucleotides which encode substantially the same or a functionally equivalent

polypeptides may be produced and used to express NAAP.

The polynucleotides of the invention can be engineered using methods generally known in the art in order to alter NAAP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of NAAP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, polynucleotides encoding NAAP may be synthesized, in whole or in part, using one or more chemical methods well known in the art (Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232). Alternatively, NAAP itself or a fragment thereof may be synthesized using chemical methods known in the art. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques (Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; Roberge, J.Y. et al. (1995) Science 269:202-204). Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of NAAP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography (Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421). The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (Creighton, *supra*, pp. 28-53).

In order to express a biologically active NAAP, the polynucleotides encoding NAAP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotides encoding NAAP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of polynucleotides encoding NAAP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where a polynucleotide sequence encoding NAAP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used (Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162).

Methods which are well known to those skilled in the art may be used to construct expression vectors containing polynucleotides encoding NAAP and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination (Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel et al., *supra*, ch. 1, 3, and 15).

A variety of expression vector/host systems may be utilized to contain and express polynucleotides encoding NAAP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems (Sambrook, *supra*; Ausubel et al., *supra*; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945;

Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355). Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of polynucleotides to the targeted organ, tissue, or cell population (Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5:350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90:6340-6344; Buller, R.M. et al. (1985) Nature 317:813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31:219-226; Verma, I.M. and N. Somia (1997) Nature 389:239-242). The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotides encoding NAAP. For example, routine cloning, subcloning, and propagation of polynucleotides encoding NAAP can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Invitrogen). Ligation of polynucleotides encoding NAAP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence (Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509). When large quantities of NAAP are needed, e.g. for the production of antibodies, vectors which direct high level expression of NAAP may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of NAAP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign polynucleotide sequences into the host genome for stable propagation (Ausubel et al., *supra*; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; Scorer, C.A. et al. (1994) Bio/Technology 12:181-184).

Plant systems may also be used for expression of NAAP. Transcription of polynucleotides encoding NAAP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated

transfection (The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196).

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, polynucleotides encoding NAAP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses NAAP in host cells (Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355).

For long term production of recombinant proteins in mammalian systems, stable expression of NAAP in cell lines is preferred. For example, polynucleotides encoding NAAP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*$^-$ and *apr*$^-$ cells, respectively (Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823). Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14). Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites (Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051). Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β-glucuronidase and its substrate β-glucuronide, or luciferase and its substrate luciferin may be used.

These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding NAAP is inserted within a marker gene sequence, transformed cells containing polynucleotides encoding NAAP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding NAAP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the polynucleotide encoding NAAP and that express NAAP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of NAAP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on NAAP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art (Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding NAAP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, polynucleotides encoding NAAP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Biosciences, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of

detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with polynucleotides encoding NAAP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode NAAP may be designed to contain signal sequences which direct secretion of NAAP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted polynucleotides or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant polynucleotides encoding NAAP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric NAAP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of NAAP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the NAAP encoding sequence and the heterologous protein sequence, so that NAAP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In another embodiment, synthesis of radiolabeled NAAP may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

NAAP, fragments of NAAP, or variants of NAAP may be used to screen for compounds that specifically bind to NAAP. One or more test compounds may be screened for specific binding to NAAP. In various embodiments, 1, 2, 3, 4, 5, 10, 20, 50, 100, or 200 test compounds can be screened for specific binding to NAAP. Examples of test compounds can include antibodies, anticalins, oligonucleotides, proteins (e.g., ligands or receptors), or small molecules.

In related embodiments, variants of NAAP can be used to screen for binding of test compounds, such as antibodies, to NAAP, a variant of NAAP, or a combination of NAAP and/or one or more variants NAAP. In an embodiment, a variant of NAAP can be used to screen for compounds that bind to a variant of NAAP, but not to NAAP having the exact sequence of a sequence of SEQ ID NO:1-35. NAAP variants used to perform such screening can have a range of about 50% to about 99% sequence identity to NAAP, with various embodiments having 60%, 70%, 75%, 80%, 85%, 90%, and 95% sequence identity.

In an embodiment, a compound identified in a screen for specific binding to NAAP can be closely related to the natural ligand of NAAP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner (Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):Chapter 5). In another embodiment, the compound thus identified can be a natural ligand of a receptor NAAP (Howard, A.D. et al. (2001) Trends Pharmacol. Sci.22:132-140; Wise, A. et al. (2002) Drug Discovery Today 7:235-246).

In other embodiments, a compound identified in a screen for specific binding to NAAP can be closely related to the natural receptor to which NAAP binds, at least a fragment of the receptor, or a fragment of the receptor including all or a portion of the ligand binding site or binding pocket. For example, the compound may be a receptor for NAAP which is capable of propagating a signal, or a decoy receptor for NAAP which is not capable of propagating a signal (Ashkenazi, A. and V.M. Divit (1999) Curr. Opin. Cell Biol. 11:255-260; Mantovani, A. et al. (2001) Trends Immunol. 22:328-336). The compound can be rationally designed using known techniques. Examples of such techniques include those used to construct the compound etanercept (ENBREL; Immunex Corp., Seattle WA), which is efficacious for treating rheumatoid arthritis in humans. Etanercept is an engineered p75 tumor necrosis factor (TNF) receptor dimer linked to the Fc portion of human IgG_1 (Taylor, P.C. et al. (2001) Curr. Opin. Immunol. 13:611-616).

In one embodiment, two or more antibodies having similar or, alternatively, different

specificities can be screened for specific binding to NAAP, fragments of NAAP, or variants of NAAP. The binding specificity of the antibodies thus screened can thereby be selected to identify particular fragments or variants of NAAP. In one embodiment, an antibody can be selected such that its binding specificity allows for preferential identification of specific fragments or variants of NAAP. In another embodiment, an antibody can be selected such that its binding specificity allows for preferential diagnosis of a specific disease or condition having increased, decreased, or otherwise abnormal production of NAAP.

In an embodiment, anticalins can be screened for specific binding to NAAP, fragments of NAAP, or variants of NAAP. Anticalins are ligand-binding proteins that have been constructed based on a lipocalin scaffold (Weiss, G.A. and H.B. Lowman (2000) Chem. Biol. 7:R177-R184; Skerra, A. (2001) J. Biotechnol. 74:257-275). The protein architecture of lipocalins can include a beta-barrel having eight antiparallel beta-strands, which supports four loops at its open end. These loops form the natural ligand-binding site of the lipocalins, a site which can be re-engineered *in vitro* by amino acid substitutions to impart novel binding specificities. The amino acid substitutions can be made using methods known in the art or described herein, and can include conservative substitutions (e.g., substitutions that do not alter binding specificity) or substitutions that modestly, moderately, or significantly alter binding specificity.

In one embodiment, screening for compounds which specifically bind to, stimulate, or inhibit NAAP involves producing appropriate cells which express NAAP, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing NAAP or cell membrane fractions which contain NAAP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either NAAP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with NAAP, either in solution or affixed to a solid support, and detecting the binding of NAAP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

An assay can be used to assess the ability of a compound to bind to its natural ligand and/or to inhibit the binding of its natural ligand to its natural receptors. Examples of such assays include radio-labeling assays such as those described in U.S. Patent No. 5,914,236 and U.S. Patent No. 6,372,724. In a related embodiment, one or more amino acid substitutions can be introduced into a

polypeptide compound (such as a receptor) to improve or alter its ability to bind to its natural ligands (Matthews, D.J. and J.A. Wells. (1994) Chem. Biol. 1:25-30). In another related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a ligand) to improve or alter its ability to bind to its natural receptors (Cunningham, B.C. and J.A. Wells (1991) Proc. Natl. Acad. Sci. USA 88:3407-3411; Lowman, H.B. et al. (1991) J. Biol. Chem. 266:10982-10988).

NAAP, fragments of NAAP, or variants of NAAP may be used to screen for compounds that modulate the activity of NAAP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for NAAP activity, wherein NAAP is combined with at least one test compound, and the activity of NAAP in the presence of a test compound is compared with the activity of NAAP in the absence of the test compound. A change in the activity of NAAP in the presence of the test compound is indicative of a compound that modulates the activity of NAAP. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising NAAP under conditions suitable for NAAP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of NAAP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding NAAP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease (see, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337). For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (*neo*; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding NAAP may also be manipulated *in vitro* in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell

lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding NAAP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding NAAP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress NAAP, e.g., by secreting NAAP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of NAAP and nucleic acid-associated proteins. In addition, examples of tissues expressing NAAP can be found in Table 6 and can also be found in Example XI. Therefore, NAAP appears to play a role in cell proliferative, neurological, developmental, and autoimmune/inflammatory disorders, and infections. In the treatment of disorders associated with increased NAAP expression or activity, it is desirable to decrease the expression or activity of NAAP. In the treatment of disorders associated with decreased NAAP expression or activity, it is desirable to increase the expression or activity of NAAP.

Therefore, in one embodiment, NAAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NAAP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial

thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorder of the central nervous system, cerebral palsy, a neuroskeletal disorder, an autonomic nervous system disorder, a cranial nerve disorder, a spinal cord disease, muscular dystrophy and other neuromuscular disorder, a peripheral nervous system disorder, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathy, myasthenia gravis, periodic paralysis, a mental disorder including mood, anxiety, and schizophrenic disorder, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia, and Tourette's disorder; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; an infection, such as those caused by a viral agent classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, hepadnavirus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, or

togavirus; an infection caused by a bacterial agent classified as pneumococcus, staphylococcus, streptococcus, bacillus, corynebacterium, clostridium, meningococcus, gonococcus, listeria, moraxella, kingella, haemophilus, legionella, bordetella, gram-negative enterobacterium including shigella, salmonella, or campylobacter, pseudomonas, vibrio, brucella, francisella, yersinia, bartonella, norcardium, actinomyces, mycobacterium, spirochaetale, rickettsia, chlamydia, or mycoplasma; an infection caused by a fungal agent classified as aspergillus, blastomyces, dermatophytes, cryptococcus, coccidioides, malasezzia, histoplasma, or other mycosis-causing fungal agent; and an infection caused by a parasite classified as plasmodium or malaria-causing, parasitic entamoeba, leishmania, trypanosoma, toxoplasma, pneumocystis carinii, intestinal protozoa such as giardia, trichomonas, tissue nematode such as trichinella, intestinal nematode such as ascaris, lymphatic filarial nematode, trematode such as schistosoma, and cestode such as tapeworm.

In another embodiment, a vector capable of expressing NAAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NAAP including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified NAAP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NAAP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of NAAP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of NAAP including, but not limited to, those listed above.

In a further embodiment, an antagonist of NAAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of NAAP. Examples of such disorders include, but are not limited to, those cell proliferative, neurological, developmental, and autoimmune/inflammatory disorders, and infections described above. In one aspect, an antibody which specifically binds NAAP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express NAAP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding NAAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of NAAP including, but not limited to, those described above.

In other embodiments, any protein, agonist, antagonist, antibody, complementary sequence, or vector embodiments may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various

disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of NAAP may be produced using methods which are generally known in the art. In particular, purified NAAP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind NAAP. Antibodies to NAAP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have advantages in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with NAAP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to NAAP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of NAAP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to NAAP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used (Morrison, S.L. et al. (1984) Proc. Natl. Acad.

Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce NAAP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137).

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

Antibody fragments which contain specific binding sites for NAAP may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W.D. et al. (1989) Science 246:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between NAAP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering NAAP epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for NAAP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of NAAP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple NAAP epitopes, represents the average affinity, or avidity, of the antibodies for NAAP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular NAAP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the NAAP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of NAAP, preferably in active form, from the

antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of NAAP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available (Catty, *supra*; Coligan et al., *supra*).

In another embodiment of the invention, polynucleotides encoding NAAP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding NAAP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding NAAP (Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press, Totawa NJ).

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein (Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102:469-475; Scanlon, K.J. et al. (1995) 9:1288-1296). Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors (Miller, A.D. (1990) Blood 76:271; Ausubel et al., *supra*; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63:323-347). Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art (Rossi, J.J. (1995) Br. Med. Bull. 51:217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87:1308-1315; Morris, M.C. et al. (1997) Nucleic Acids Res. 25:2730-2736).

In another embodiment of the invention, polynucleotides encoding NAAP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene

Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides brasiliensis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the case where a genetic deficiency in NAAP expression or regulation causes disease, the expression of NAAP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in NAAP are treated by constructing mammalian expression vectors encoding NAAP and introducing these vectors by mechanical means into NAAP-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J.-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of NAAP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). NAAP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, *supra*)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding NAAP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver

polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to NAAP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding NAAP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In an embodiment, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding NAAP to cells which have one or more genetic abnormalities with respect to the expression of NAAP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999; Annu. Rev. Nutr. 19:511-544) and Verma, I.M. and N. Somia (1997; Nature 18:389:239-242).

In another embodiment, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding NAAP to target cells which have one or more genetic abnormalities with respect to the expression of NAAP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing NAAP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999; J. Virol. 73:519-532) and Xu, H. et al. (1994; Dev. Biol. 163:152-161). The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another embodiment, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding NAAP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for NAAP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of NAAP-coding RNAs and the synthesis of high levels of NAAP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of NAAP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA

transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177). A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of RNA molecules encoding NAAP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA molecules encoding NAAP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine,

queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding NAAP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased NAAP expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding NAAP may be therapeutically useful, and in the treatment of disorders associated with decreased NAAP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding NAAP may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding NAAP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding NAAP are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding NAAP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun.

268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art (Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466).

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of NAAP, antibodies to NAAP, and mimetics, agonists, antagonists, or inhibitors of NAAP.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of

macromolecules comprising NAAP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, NAAP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example NAAP or fragments thereof, antibodies of NAAP, and agonists, antagonists or inhibitors of NAAP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells,

conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind NAAP may be used for the diagnosis of disorders characterized by expression of NAAP, or in assays to monitor patients being treated with NAAP or agonists, antagonists, or inhibitors of NAAP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for NAAP include methods which utilize the antibody and a label to detect NAAP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring NAAP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of NAAP expression. Normal or standard values for NAAP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to NAAP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of NAAP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, polynucleotides encoding NAAP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotides, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of NAAP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of NAAP, and to monitor regulation of NAAP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotides, including genomic sequences, encoding NAAP or closely related molecules may be used to identify nucleic acid sequences which encode NAAP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding NAAP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the NAAP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:36-70 or from genomic sequences including promoters, enhancers, and introns of the NAAP gene.

Means for producing specific hybridization probes for polynucleotides encoding NAAP

include the cloning of polynucleotides encoding NAAP or NAAP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotides encoding NAAP may be used for the diagnosis of disorders associated with expression of NAAP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorder of the central nervous system, cerebral palsy, a neuroskeletal disorder, an autonomic nervous system disorder, a cranial nerve disorder, a spinal cord disease, muscular dystrophy and other neuromuscular disorder, a peripheral nervous system disorder, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathy, myasthenia gravis, periodic paralysis, a mental disorder including mood, anxiety, and schizophrenic disorder, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary

keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; an infection, such as those caused by a viral agent classified as adenovirus, arenavirus, bunyavirus, calicivirus, coronavirus, filovirus, hepadnavirus, herpesvirus, flavivirus, orthomyxovirus, parvovirus, papovavirus, paramyxovirus, picornavirus, poxvirus, reovirus, retrovirus, rhabdovirus, or togavirus; an infection caused by a bacterial agent classified as pneumococcus, staphylococcus, streptococcus, bacillus, corynebacterium, clostridium, meningococcus, gonococcus, listeria, moraxella, kingella, haemophilus, legionella, bordetella, gram-negative enterobacterium including shigella, salmonella, or campylobacter, pseudomonas, vibrio, brucella, francisella, yersinia, bartonella, norcardium, actinomyces, mycobacterium, spirochaetale, rickettsia, chlamydia, or mycoplasma; an infection caused by a fungal agent classified as aspergillus, blastomyces, dermatophytes, cryptococcus, coccidioides, malasezzia, histoplasma, or other mycosis-causing fungal agent; and an infection caused by a parasite classified as plasmodium or malaria-causing, parasitic entamoeba, leishmania, trypanosoma, toxoplasma, pneumocystis carinii, intestinal protozoa such as giardia, trichomonas, tissue nematode such as trichinella, intestinal nematode such as ascaris, lymphatic filarial nematode, trematode such as schistosoma, and cestode such as tapeworm. Polynucleotides encoding NAAP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered NAAP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, polynucleotides encoding NAAP may be used in assays that detect the

presence of associated disorders, particularly those mentioned above. Polynucleotides complementary to sequences encoding NAAP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of polynucleotides encoding NAAP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of NAAP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding NAAP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier, thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding NAAP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding NAAP, or a fragment of a polynucleotide complementary to the polynucleotide encoding NAAP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or

quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from polynucleotides encoding NAAP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from polynucleotides encoding NAAP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641).

Methods which may also be used to quantify the expression of NAAP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from

standard curves (Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotides described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, NAAP, fragments of NAAP, or antibodies specific for NAAP may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time (Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484; hereby expressly incorporated by reference herein). Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression *in vivo*, as in the case of a tissue or biopsy sample, or *in vitro*, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present

invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity (see, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm). Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In an embodiment, the toxicity of a test compound can be assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another embodiment relates to the use of the polypeptides disclosed herein to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are

visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of interest. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for NAAP to quantify the levels of NAAP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological

sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art (Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662). Various types of microarrays are well known and thoroughly described in Schena, M., ed. (1999; DNA Microarrays: A Practical Approach, Oxford University Press, London).

In another embodiment of the invention, nucleic acid sequences encoding NAAP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; Trask, B.J. (1991) Trends Genet. 7:149-154). Once mapped, the nucleic acid sequences may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP) (Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357).

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data (Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968). Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding NAAP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps.

Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation (Gatti, R.A. et al. (1988) Nature 336:577-580). The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, NAAP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between NAAP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest (Geysen, et al. (1984) PCT application WO84/03564). In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with NAAP, or fragments thereof, and washed. Bound NAAP is then detected by methods well known in the art. Purified NAAP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding NAAP specifically compete with a test compound for binding NAAP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with NAAP.

In additional embodiments, the nucleotide sequences which encode NAAP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, including U.S. Ser. No. 60/305,089, U.S. Ser. No. 60/305,104, U.S. Ser. No. 60/305,325, U.S. Ser.

No. 60/305,390, U.S. Ser. No. 60/306,694, U.S. Ser. No. 60/306,960, and U.S. Ser. No. 60/308,170, are expressly incorporated by reference herein.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Invitrogen), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Invitrogen), using the recommended procedures or similar methods known in the art (Ausubel et al., *supra*, ch. 5). Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Biosciences) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Invitrogen), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Invitrogen.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by *in vivo*

excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Biosciences or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Amersham Biosciences); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (Ausubel et al., *supra*, ch. 7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Candida albicans* (Incyte Genomics, Palo Alto CA); hidden

Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as SMART (Schultz, J. et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002) Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families; see, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:36-70. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative nucleic acid-associated proteins were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94; Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode nucleic acid-associated proteins, the encoded polypeptides were analyzed by querying against PFAM models for nucleic acid-associated proteins. Potential nucleic acid-associated proteins were also identified by homology to Incyte cDNA sequences that had been annotated as nucleic acid-associated proteins. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated

but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of NAAP Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:36-70 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:36-70 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between

chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook, *supra*, ch. 7; Ausubel et al., *supra*, ch. 4).

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum}\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotides encoding NAAP are analyzed with respect to the tissue sources

from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding NAAP. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of NAAP Encoding Polynucleotides

Full length polynucleotides are produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Biosciences), ELONGASE enzyme (Invitrogen), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 μl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Biosciences). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Biosciences), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Biosciences) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Biosciences) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotides are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Identification of Single Nucleotide Polymorphisms in NAAP Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:36-70 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters was

used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezualan, and two Amish individuals. The African population comprised 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed no allelic variance in this population were not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:36-70 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Biosciences), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Biosciences). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature

under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing; see, e.g., Baldeschweiler et al., *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena, M., ed. (1999) DNA Microarrays: A Practical Approach, Oxford University Press, London). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements (Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31).

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Biosciences). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by *in vitro* transcription from non-coding

yeast genomic DNA. After incubation at 37° C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85° C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Biosciences).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60° C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65° C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of

140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then

integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

Expression

Array elements that exhibited at least about a two-fold change in expression, a signal-to-background ratio of at least 2.5, and an element spot size of at least 40% were identified as differentially expressed using the GEMTOOLS program (Incyte Genomics).

SEQ ID NO:57 showed differential expression, as determined by microarray analysis, in human aortic endothelial cells (HAEC) following exposure to 10 ng/ml TNF-a for 24 and 48 hours. TNF-a is a pleiotropic cytokine that is known to play a central role in the mediation of inflammatory responses through activation of multiple signal transduction pathways. HAECs are primary cells derived from the endothelium of a human aorta. These cells were grown to 85% confluency and then treated with TNF-a. The expression of SEQ ID NO:57 was increased by at least two-fold in TNF-a-treated HAECs, as compared to untreated controls. Therefore, in various embodiments, SEQ ID NO:57 can be used for one or more of the following: i) monitoring treatment of immune disorders and related diseases and conditions, ii) diagnostic assays for immune disorders and related diseases and conditions, and iii) developing therapeutics and/or other treatments for immune disorders and related diseases and conditions.

XII. Complementary Polynucleotides

Sequences complementary to the NAAP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring NAAP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of NAAP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the NAAP-encoding transcript.

XIII. Expression of NAAP

Expression and purification of NAAP is achieved using bacterial or virus-based expression systems. For expression of NAAP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express NAAP upon induction with isopropyl beta-D-

thiogalactopyranoside (IPTG). Expression of NAAP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding NAAP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus (Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945).

In most expression systems, NAAP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Biosciences). Following purification, the GST moiety can be proteolytically cleaved from NAAP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). Purified NAAP obtained by these methods can be used directly in the assays shown in Examples XVII, XVIII, and XIX, where applicable.

XIV. Functional Assays

NAAP function is assessed by expressing the sequences encoding NAAP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT plasmid (Invitrogen, Carlsbad CA) and PCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to

evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994; Flow Cytometry, Oxford, New York NY).

The influence of NAAP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding NAAP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding NAAP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of NAAP Specific Antibodies

NAAP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the NAAP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art (Ausubel et al., *supra*, ch. 11).

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity (Ausubel et al., *supra*). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-NAAP activity by, for example, binding the peptide or NAAP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XVI. Purification of Naturally Occurring NAAP Using Specific Antibodies

Naturally occurring or recombinant NAAP is substantially purified by immunoaffinity

chromatography using antibodies specific for NAAP. An immunoaffinity column is constructed by covalently coupling anti-NAAP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Biosciences). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing NAAP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of NAAP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/NAAP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and NAAP is collected.

XVII. Identification of Molecules Which Interact with NAAP

NAAP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent (Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539). Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled NAAP, washed, and any wells with labeled NAAP complex are assayed. Data obtained using different concentrations of NAAP are used to calculate values for the number, affinity, and association of NAAP with the candidate molecules.

Alternatively, molecules interacting with NAAP are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989; Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

NAAP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVIII. Demonstration of NAAP Activity

NAAP activity is measured by its ability to stimulate transcription of a reporter gene (Liu, H.Y. et al. (1997) EMBO J. 16:5289-5298). The assay entails the use of a well characterized reporter gene construct, $LexA_{op}$-LacZ, that consists of LexA DNA transcriptional control elements ($LexA_{op}$) fused to sequences encoding the E. coli LacZ enzyme. The methods for constructing and expressing fusion genes, introducing them into cells, and measuring LacZ enzyme activity, are well known to those skilled in the art. Sequences encoding NAAP are cloned into a plasmid that directs the synthesis of a fusion protein, LexA-NAAP, consisting of NAAP and a DNA binding domain derived from the LexA transcription factor. The resulting plasmid, encoding a LexA-NAAP fusion protein, is introduced into yeast cells along with a plasmid containing the $LexA_{op}$-LacZ reporter gene. The amount of LacZ enzyme activity associated with LexA-NAAP transfected cells, relative to control cells, is proportional to the amount of transcription stimulated by the NAAP.

Alternatively, NAAP activity is measured by its ability to bind zinc. A 5-10 mM sample solution in 2.5 mM ammonium acetate solution at pH 7.4 is combined with 0.05 M zinc sulfate solution (Aldrich, Milwaukee WI) in the presence of 100 mM dithiothreitol with 10% methanol added. The sample and zinc sulfate solutions are allowed to incubate for 20 minutes. The reaction solution is passed through a VYDAC column (Grace Vydac, Hesperia, CA) with approximately 300 Angstrom bore size and 5 mM particle size to isolate zinc-sample complex from the solution, and into a mass spectrometer (PE Sciex, Ontario, Canada). Zinc bound to sample is quantified using the functional atomic mass of 63.5 Da observed by Whittal, R.M. et al. ((2000) Biochemistry 39:8406-8417).

In the alternative, a method to determine nucleic acid binding activity of NAAP involves a polyacrylamide gel mobility-shift assay. In preparation for this assay, NAAP is expressed by transforming a mammalian cell line such as COS7, HeLa or CHO with a eukaryotic expression vector containing NAAP cDNA. The cells are incubated for 48-72 hours after transformation under conditions appropriate for the cell line to allow expression and accumulation of NAAP. Extracts containing solubilized proteins can be prepared from cells expressing NAAP by methods well known in the art. Portions of the extract containing NAAP are added to [^{32}P]-labeled RNA or DNA. Radioactive nucleic acid can be synthesized *in vitro* by techniques well known in the art. The mixtures are incubated at 25°C in the presence of RNase- and DNase-inhibitors under buffered conditions for 5-10 minutes. After incubation, the samples are analyzed by polyacrylamide gel electrophoresis followed by autoradiography. The presence of a band on the autoradiogram indicates the formation of a complex between NAAP and the radioactive transcript. A band of similar mobility will not be present in samples prepared using control extracts prepared from untransformed cells.

In the alternative, a method to determine methylase activity of NAAP measures transfer of radiolabeled methyl groups between a donor substrate and an acceptor substrate. Reaction mixtures (50 μl final volume) contain 15 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 10 mM dithiothreitol, 3% polyvinylalcohol, 1.5 μCi [*methyl*-^{3}H]AdoMet (0.375 μM AdoMet) (DuPont-NEN), 0.6 μg NAAP, and acceptor substrate (e.g., 0.4 μg [^{35}S]RNA, or 6-mercaptopurine (6-MP) to 1 mM final concentration). Reaction mixtures are incubated at 30°C for 30 minutes, then 65°C for 5 minutes.

Analysis of [*methyl*-^{3}H]RNA is as follows: (1) 50 μl of 2 x loading buffer (20 mM Tris-HCl, pH 7.6, 1 M LiCl, 1 mM EDTA, 1% sodium dodecyl sulphate (SDS)) and 50 μl oligo d(T)-cellulose (10 mg/ml in 1 x loading buffer) are added to the reaction mixture, and incubated at ambient temperature with shaking for 30 minutes. (2) Reaction mixtures are transferred to a 96-well filtration plate attached to a vacuum apparatus. (3) Each sample is washed sequentially with three 2.4 ml aliquots of 1 x oligo d(T) loading buffer containing 0.5% SDS, 0.1% SDS, or no SDS. (4) RNA is eluted with 300 μl of water into a 96-well collection plate, transferred to scintillation vials containing

liquid scintillant, and radioactivity determined.

Analysis of [*methyl*-^{3}H]6-MP is as follows: (1) 500 µl 0.5 M borate buffer, pH 10.0, and then 2.5 ml of 20% (v/v) isoamyl alcohol in toluene are added to the reaction mixtures. (2) The samples are mixed by vigorous vortexing for ten seconds. (3) After centrifugation at 700g for 10 minutes, 1.5 ml of the organic phase is transferred to scintillation vials containing 0.5 ml absolute ethanol and liquid scintillant, and radioactivity determined. (4) Results are corrected for the extraction of 6-MP into the organic phase (approximately 41%).

In the alternative, type I topoisomerase activity of NAAP can be assayed based on the relaxation of a supercoiled DNA substrate. NAAP is incubated with its substrate in a buffer lacking Mg^{2+} and ATP, the reaction is terminated, and the products are loaded on an agarose gel. Altered topoisomers can be distinguished from supercoiled substrate electrophoretically. This assay is specific for type I topoisomerase activity because Mg^{2+} and ATP are necessary cofactors for type II topoisomerases.

Type II topoisomerase activity of NAAP can be assayed based on the decatenation of a kinetoplast DNA (KDNA) substrate. NAAP is incubated with KDNA, the reaction is terminated, and the products are loaded on an agarose gel. Monomeric circular KDNA can be distinguished from catenated KDNA electrophoretically. Kits for measuring type I and type II topoisomerase activities are available commercially from Topogen (Columbus OH).

ATP-dependent RNA helicase unwinding activity of NAAP can be measured by the method described by Zhang and Grosse (1994; Biochemistry 33:3906-3912). The substrate for RNA unwinding consists of ^{32}P-labeled RNA composed of two RNA strands of 194 and 130 nucleotides in length containing a duplex region of 17 base-pairs. The RNA substrate is incubated together with ATP, Mg^{2+}, and varying amounts of NAAP in a Tris-HCl buffer, pH 7.5, at 37°C for 30 minutes. The single-stranded RNA product is then separated from the double-stranded RNA substrate by electrophoresis through a 10% SDS-polyacrylamide gel, and quantitated by autoradiography. The amount of single-stranded RNA recovered is proportional to the amount of NAAP in the preparation.

Splicing activity of NAAP can be measured by the method of Hartmann, A.M. et al. (*supra*). Varying amounts of a construct containing NAAP, for example cloned into an expression vector such as PEGFP-C2 (Clontech), are transfected into HEK293 cells using the calcium phosphate method as described. RNA is isolated 17-24 hours after the transfection using the RNEASY mini kit (QIAGEN). Isolated RNA is mixed with antisense primer and dNTP and subjected to reaction with reverse transcriptase. Products of the reverse transcriptase reaction are amplified by PCR and analyzed on a 2% agarose Tris borate-EDTA gel.

In the alternative, NAAP function is assessed by expressing the sequences encoding NAAP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a

mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT (Life Technologies) and pCR3.1 (Invitrogen Corporation, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, preferably of endothelial or hematopoietic origin, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected.

Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; CLONTECH), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties.

FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M. G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of NAAP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding NAAP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Inc., Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding NAAP and other genes of interest can be analyzed by northern analysis or microarray techniques.

Pseudouridine synthase activity of NAAP is assayed using a tritium (^3H) release assay modified from Nurse et al. ((1995) RNA 1:102-112), which measures the release of ^3H from the C_5 position of the pyrimidine component of uridylate (U) when ^3H-radiolabeled U in RNA is isomerized to pseudouridine (y). A typical 500 μl assay mixture contains 50 mM HEPES buffer (pH 7.5), 100 mM ammonium acetate, 5 mM dithiothreitol, 1 mM EDTA, 30 units RNase inhibitor, and 0.1-4.2 μM [5-^3H]tRNA (approximately 1 μCi/nmol tRNA). The reaction is initiated by the addition of <5 μl of a

concentrated solution of NAAP (or sample containing NAAP) and incubated for 5 min at 37 °C. Portions of the reaction mixture are removed at various times (up to 30 min) following the addition of NAAP and quenched by dilution into 1 ml 0.1 M HCl containing Norit-SA3 (12% w/v). The quenched reaction mixtures are centrifuged for 5 min at maximum speed in a microcentrifuge, and the supernatants are filtered through a plug of glass wool. The pellet is washed twice by resuspension in 1 ml 0.1 M HCl, followed by centrifugation. The supernatants from the washes are separately passed through the glass wool plug and combined with the original filtrate. A portion of the combined filtrate is mixed with scintillation fluid (up to 10 ml) and counted using a scintillation counter. The amount of ^{3}H released from the RNA and present in the soluble filtrate is proportional to the amount of peudouridine synthase activity in the sample (Ramamurthy, V. (1999) J. Biol. Chem. 274:22225-22230).

In the alternative, pseudouridine synthase activity of NAAP is assayed at 30 °C to 37 °C in a mixture containing 100 mM Tris-HCl (pH 8.0), 100 mM ammonium acetate, 5 mM $MgCl_2$, 2 mM dithiothreitol, 0.1 mM EDTA, and 1-2 fmol of [^{32}P]-radiolabeled runoff transcripts (generated *in vitro* by an appropriate RNA polymerase, i.e., T7 or SP6) as substrates. NAAP is added to initiate the reaction or omitted from the reaction in control samples. Following incubation, the RNA is extracted with phenol-chloroform, precipitated in ethanol, and hydrolyzed completely to 3-nucleotide monophosphates using RNase T_2. The hydrolysates are analyzed by two-dimensional thin layer chromatography, and the amount of ^{32}P radiolabel present in the yMP and UMP spots are evaluated after exposing the thin layer chromatography plates to film or a PhosphorImager screen. Taking into account the relative number of uridylate residues in the substrate RNA, the relative amount yMP and UMP are determined and used to calculate the relative amount of y per tRNA molecule (expressed in mol y /mol of tRNA or mol y /mol of tRNA/minute), which corresponds to the amount of pseudouridine synthase activity in the NAAP sample (Lecointe, *supra*).

N^2,N^2-dimethylguanosine transferase ((m^2_2G)methyltransferase) activity of NAAP is measured in a 160 μl reaction mixture containing 100 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 10 mM $MgCl_2$, 20 mM NH_4Cl, 1mM dithiothreitol, 6.2 μM *S*-adenosyl-L-[*methyl*-^{3}H]methionine (30-70 Ci/mM), 8 μg m^2_2G-deficient tRNA or wild type tRNA from yeast, and approximately 100 μg of purified NAAP or a sample comprising NAAP. The reactions are incubated at 30 °C for 90 min and chilled on ice. A portion of each reaction is diluted to 1 ml in water containing 100 μg BSA. 1 ml of 2 M HCl is added to each sample and the acid insoluble products are allowed to precipitate on ice for 20 min before being collected by filtration through glass fiber filters. The collected material is washed several times with HCl and quantitated using a liquid scintillation counter. The amount of ^{3}H incorporated into the m^2_2G-deficient, acid-insoluble tRNAs is proportional to the amount of N^2,N^2-dimethylguanosine transferase activity in the NAAP sample. Reactions comprising no

substrate tRNAs, or wild-type tRNAs that have already been modified, serve as control reactions which should not yield acid-insoluble ^{3}H-labeled products.

Polyadenylation activity of NAAP is measured using an *in vitro* polyadenylation reaction. The reaction mixture is assembled on ice and comprises 10 μl of 5 mM dithiothreitol, 0.025% (v/v) NONIDET P-40, 50 mM creatine phosphate, 6.5% (w/v) polyvinyl alcohol, 0.5 unit/μl RNAGUARD (Pharmacia), 0.025 μg/μl creatine kinase, 1.25 mM cordycepin 5'-triphosphate, and 3.75 mM $MgCl_2$, in a total volume of 25 μl. 60 fmol of CstF, 50 fmol of CPSF, 240 fmol of PAP, 4 μl of crude or partially purified CF II and various amounts of amounts CF I are then added to the reaction mix. The volume is adjusted to 23.5 μl with a buffer containing 50 mM TrisHCl, pH 7.9, 10% (v/v) glycerol, and 0.1 mM Na-EDTA. The final ammonium sulfate concentration should be below 20 mM. The reaction is initiated (on ice) by the addition of 15 fmol of ^{32}P-labeled pre-mRNA template, along with 2.5 μg of unlabeled tRNA, in 1.5 μl of water. Reactions are then incubated at 30 °C for 75-90 min and stopped by the addition of 75 μl (approximately two-volumes) of proteinase K mix (0.2 M Tris-HCl, pH 7.9, 300 mM NaCl, 25 mM Na-EDTA, 2% (w/v) SDS), 1 μl of 10 mg/ml proteinase K, 0.25 μl of 20 mg/ml glycogen, and 23.75 μl of water). Following incubation, the RNA is precipitated with ethanol and analyzed on a 6% (w/v) polyacrylamide, 8.3 M urea sequencing gel. The dried gel is developed by autoradiography or using a phosphoimager. Cleavage activity is determined by comparing the amount of cleavage product to the amount of pre-mRNA template. The omission of any of the polypeptide components of the reaction and substitution of NAAP is useful for identifying the specific biological function of NAAP in pre-mRNA polyadenylation (Rüegsegger, *supra*; and references within).

tRNA synthetase activity is measured as the aminoacylation of a substrate tRNA in the presence of [^{14}C]-labeled amino acid. NAAP is incubated with [^{14}C]-labeled amino acid and the appropriate cognate tRNA (for example, [^{14}C]alanine and $tRNA^{ala}$) in a buffered solution. ^{14}C-labeled product is separated from free [^{14}C]amino acid by chromatography, and the incorporated ^{14}C is quantified by scintillation counter. The amount of ^{14}C-labeled product detected is proportional to the activity of NAAP in this assay.

In the alternative, NAAP activity is measured by incubating a sample containing NAAP in a solution containing 1 mM ATP, 5 mM Hepes-KOH (pH 7.0), 2.5 mM KCl, 1.5 mM magnesium chloride, and 0.5 mM DTT along with misacylated [^{14}C]-Glu-tRNAGln (e.g., 1 μM) and a similar concentration of unlabeled L-glutamine. Following the quenching of the reaction with 3 M sodium acetate (pH 5.0), the mixture is extracted with an equal volume of water-saturated phenol, and the aqueous and organic phases are separated by centrifugation at 15,000 × g at room temperature for 1 min. The aqueous phase is removed and precipitated with 3 volumes of ethanol at -70°C for 15 min. The precipitated aminoacyl-tRNAs are recovered by centrifugation at 15,000 × g at 4°C for15 min.

The pellet is resuspended in of 25 mM KOH, deacylated at 65°C for 10 min., neutralized with 0.1 M HCl (to final pH 6-7), and dried under vacuum. The dried pellet is resuspended in water and spotted onto a cellulose TLC plate. The plate is developed in either isopropanol/formic acid/water or ammonia/water/chloroform/ methanol. The image is subjected to densitometric analysis and the relative amounts of Glu and Gln are calculated based on the Rf values and relative intensities of the spots. NAAP activity is calculated based on the amount of Gln resulting from the transformation of Glu while acylated as Glu-tRNA^{Gln} (adapted from Curnow, A.W. et al. (1997) Proc. Natl. Acad. Sci. USA 94:11819-26).

An alternative experiment for NAAP activity involves binding of DNA-bound KAP-1-RBBC protein, a corepressor for KRAB domain proteins, directly to the KRAB domain. Following preparation of plasmids and protein purification (Peng, H., et al. (2000) J. Biol. Chem. 275:18000-18010), an electrophilic mobility shift assay (EMSA) can be performed in which purified recombinant GAL4-KRAB protein is incubated with purified Escherichia coli- or baculovirus-expressed KAP-1-RBBC protein for 15 min at 30°C. The KRAB protein is then added to the reaction simultaneously with the GAL4-KRAB and KAP-1-RBBC proteins, or the KRAB protein can be pre-incubated with the KAP-1-RBBC protein for 15 min at 30°C. One ml of ^{32}P-labeled GAL4 probe (10^5 cpm/ml) is then added, and the reaction incubated for an additional 15 min at 30°C. The DNA-protein complexes are then resolved on native polyacrylamide gels by electrophoresis in 45 mM Tris borate, pH 8.3, 1 mM EDTA buffer at 4°C. The EMSA gels are dried and visualized by autoradiography. Binding of the GAL4-KRAB protein complex to a standard ^{32}P-labeled GAL4 oligonucleotide recognition sequence is demonstration of a mobility shift, and indicative of KRAB domain binding via direct interaction between the KRAB domain and KAP-1 protein.

NAAP activity can be demonstrated by the use of *in vitro* translation assays which utilize mutant strains of S. cervisiae lacking the *FUN12* gene which encodes yeast translation initiation factor 2 (IF2). These strains exhibit a slow growth phenotype which can be rescued (made to grow at a normal rate) by the addition of IF2, including heterologous IF2 which is produced by recombinant methods. Briefly, the *fun 12Δ* strain J133 is transformed with either the low copy-number *FUN12* plasmid pC479, an expression plasmid carrying NAAP, or the vector only. The control strains and the test strains are streaked on synthetic minimal medium containing 10% galactose plus the required nutrient supplements, and the plates are incubated at 30°C for 5 days. In vitro translation extracts are prepared from the *fun 12Δ* strain J133. Extracts are incubated with 200 ng of luciferase mRNA and increasing amounts of the control strains or the test strains containing a source of IF2. Luminescence of the samples is plotted as a function of the amount of test protein added to the translation reaction. The amount of luminescence corresponds to the amount of NAAP activity in the sample (Lee et al. supra).

XIX. Identification of NAAP Agonists and Antagonists

Agonists or antagonists of NAAP activation or inhibition may be tested using the assays described in section XVIII. Agonists cause an increase in NAAP activity and antagonists cause a decrease in NAAP activity.

Various modifications and variations of the described compositions, methods, and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. It will be appreciated that the invention provides novel and useful proteins, and their encoding polynucleotides, which can be used in the drug discovery process, as well as methods for using these compositions for the detection, diagnosis, and treatment of diseases and conditions. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Nor should the description of such embodiments be considered exhaustive or limit the invention to the precise forms disclosed. Furthermore, elements from one embodiment can be readily recombined with elements from one or more other embodiments. Such combinations can form a number of embodiments within the scope of the invention. It is intended that the scope of the invention be defined by the following claims and their equivalents.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
7492673	1	7492673CD1	36	7492673CB1	
7990930	2	7990930CD1	37	7990930CB1	
7037554	3	7037554CD1	38	7037554CB1	
1515347	4	1515347CD1	39	1515347CB1	
3464492	5	3464492CD1	40	3464492CB1	
1794336	6	1794336CD1	41	1794336CB1	3747831CA2, 90160123CA2
2903694	7	2903694CD1	42	2903694CB1	
6975426	8	6975426CD1	43	6975426CB1	
4019390	9	4019390CD1	44	4019390CB1	
986452	10	986452CD1	45	986452CB1	
2807579	11	2807579CD1	46	2807579CB1	
5724273	12	5724273CD1	47	5724273CB1	6488912CA2
3614884	13	3614884CD1	48	3614884CB1	2889611CA2
3794954	14	3794954CD1	49	3794954CB1	
7399016	15	7399016CD1	50	7399016CB1	
6996690	16	6996690CD1	51	6996690CB1	
7740866	17	7740866CD1	52	7740866CB1	
8181605	18	8181605CD1	53	8181605CB1	
8266487	19	8266487CD1	54	8266487CB1	
5552784	20	5552784CD1	55	5552784CB1	
7281230	21	7281230CD1	56	7281230CB1	
7488424	22	7488424CD1	57	7488424CB1	2435446CA2
7487110	23	7487110CD1	58	7487110CB1	
7495008	24	7495008CD1	59	7495008CB1	
7073515	25	7073515CD1	60	7073515CB1	
3356640	26	3356640CD1	61	3356640CB1	
2015706	27	2015706CD1	62	2015706CB1	2170810CA2

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
6920755	28	6920755CD1	63	6920755CB1	90131785CA2, 90131837CA2, 90131896CA2, 90132093CA2
444179	29	444179CD1	64	444179CB1	90185918CA2
5628380	30	5628380CD1	65	5628380CB1	
7493789	31	7493789CD1	66	7493789CB1	
2075194	32	2075194CD1	67	2075194CB1	
2801633	33	2801633CD1	68	2801633CB1	3174556CA2
7493525	34	7493525CD1	69	7493525CB1	90127510CA2, 90127526CA2, 90127542CA2, 90127602CA2, 90127626CA2, 90127634CA2, 90127642CA2, 90188617CA2, 90188633CA2, 90188641CA2, 90188650CA2, 90188701CA2, 90188725CA2, 90188741CA2, 90189241CA2, 90191237CA2
7021892	35	7021892CD1	70	7021892CB1	

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	7492673CD1	g57718	2.20E-116	[Rattus rattus] ribosomal protein S2 Suzuki,K. et al. Primary structure of rat ribosomal protein S2. A ribosomal protein with arginine-glycine tandem repeats and RGGF motifs that are associated with nucleolar localization and binding to ribonucleic acids J. Biol. Chem. 266, 20007-20010 (1991)
2	7990930CD1	g1323733	6.10E-73	[Homo sapiens] ribosomal protein L9 Mazuruk,K. et al. Structural organization and chromosomal localization of the human ribosomal protein L9 gene Biochim. Biophys. Acta 1305 (3), 151-162 (1996)
3	7037554CD1	g5917651	0.0	[Rattus norvegicus] putative splicing factor YT521-B Hartmann,A.M. et al. The interaction and colocalization of Sam68 with the splicing-associated factor YT521-B in nuclear dots is regulated by the Src family kinase p59fyn Mol. Biol. Cell 10, 3909-3926 (1999)
4	1515347CD1	g8953897	6.80E-137	[Drosophila melanogaster] helicase DOMINO A
5	3464492CD1	g19110784	1.00E-144	[5' incom][Mus musculus] DNA helicase HEL308 Marini,F. and Wood,R.D. (2002) J. Biol. Chem. 277:8716-8723 A Human DNA Helicase Homologous to the DNA Cross-link Sensitivity Protein Mus308
6	1794336CD1	g6467200	2.70E-189	[Homo sapiens] gonadotropin inducible transcription repressor-1
7	2903694CD1	g7381239	1.00E-160	[fl][Mus musculus] p38 interacting protein
8	6975426CD1	g556219	0.0	[Mus musculus] transcription regulator Halleck, M.S. et al. (1995) Genomics 26:403-406 A widely distributed putative mammalian transcriptional regulator containing multiple paired amphipathic helices, with similarity to yeast SIN3
9	4019390CD1	g1017722	1.30E-157	[Homo sapiens] repressor transcriptional factor

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
10	986452CD1	g162983	6.70E-115	[Bos taurus] transcription factor EF1(A) Ozer, J. et al. (1993) Gene 124: 223-230 Isolation of the CCAAT transcription factor subunit EFIA cDNA and a potentially functional EFIA processed pseudogene from Bos taurus: insights into the evolution of the EFIA/dbpB/YB-1 gene family
11	2807579CD1	g11118640	2.80E-93	[Mus musculus] fork head transcription factor Fhx Granadino, B. et al. (2000) Mech. Dev. 97:157-160 Fhx (Foxj2) expression is activated during spermatogenesis and very early in embryonic development
12	5724273CD1	g10048470	3.00E-17	[Homo sapiens] C2H2-like zinc finger protein (Zhang, W.H., et al. (2001) 1518:190-193)
13	3614884CD1	g6693371	1.80E-231	[Homo sapiens] ZNF225
14	3794954CD1	g200407	5.10E-37	[Mus musculus] pMLZ-4 (Brady, J.P. and Piatigorsky, J. (1993) Gene 124:207-214)
15	7399016CD1	g9886891	9.80E-169	[Mus musculus] zinc finger protein 276 C2H2 type (Wong, J.C., et al. (2000) Genomics 67:273-283)
16	6996690CD1	g498721	1.10E-105	[Homo sapiens] zinc finger protein (Abrink, M., et al. (1995) DNA Cell Biol. 14:125-136)
17	7740866CD1	g186774	1.70E-212	[Homo sapiens] zinc finger protein (Bellefroid, E.J., et al. (1991) Proc. Natl. Acad. Sci. USA 88:3608-3612)
18	8181605CD1	g3970712	1.00E-28	[Homo sapiens] zinc finger protein 10 (Thiesen, H.J. (1990) New Biol. 2:363-374)
19	8266487CD1	g292931	5.40E-152	[Homo sapiens] DNA-binding protein (Greig, G.M., et al. (1993) Hum. Mol. Genet. 2:1611-1618)
20	5552784CD1	g1086577	1.10E-41	[Xenopus laevis] xbmi-1 (Reijnen, M.J., et al. (1995) Mech. Dev. 53:35-46)
21	7281230CD1	g11527849	6.20E-133	[Mus musculus] zinc finger protein SKAT2 (Blanchard, A.D., et al. (2000) Eur. J. Immunol. 30:3100-3110)
22	7488424CD1	g9650982	1.80E-215	[Homo sapiens] testis-specific RING Finger protein (Yoshikawa, T., et al. (2000) Biochim. Biophys. Acta 1493:349-355)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
23	7487110CD1	g3452555	3.30E-193	[Rattus norvegicus] BarH-class homeodomain transcription factor (Saito,T. et al. (1998) Dev. Biol. 199 (2), 216-225)
24	7495008CD1	g7385152	6.60E-108	[Mus musculus] oligodendrocyte-specific bHLH transcription factor Olig1 (Zhou,Q. et al. (2000) Neuron 25 (2), 331-343)
25	7073515CD1	g3641258	2.70E-166	[Mus musculus] ventral anterior homeobox-containing protein 1 (Hallonet,M. et al. (1998) Development 125 (14), 2599-2610)
26	3356640CD1	g1017722	1.20E-106	[Homo sapiens] repressor transcriptional factor
27	2015706CD1	g1017722	1.40E-135	[Homo sapiens] repressor transcriptional factor
28	6920755CD1	g202271	1.60E-72	[Mus musculus] UCR-motif DNA binding protein (Flanagan,J.R. et al. (1992) Mol. Cell. Biol. 12, 38-44)
29	444179CD1	g2306773	1.80E-98	[Homo sapiens] zinc finger protein (Lee,P.L. et al. (1997) Genomics 43 (2), 191-201)
30	5628380CD1	g6941960	9.10E-132	[Homo sapiens] LBP-32
31	7493789CD1	g6002623	0.0	[Homo sapiens] putative RNA-binding protein Q99 Koesters, R. et al. (1999) Human eukaryotic initiation factor EIF2C1 gene: cDNA sequence, genomic organization, localization to chromosomal bands 1p34-p35, and expression. Genomics. 61:210-218.
32	2075194CD1	g868160	3.70E-57	[Rattus norvegicus] Cys2/His2 zinc finger protein (Pott,U. et al. (1995) J. Neurochem. 65 (5), 1955-1966)
33	2801633CD1	g347906	3.60E-100	[Homo sapiens] zinc finger protein (Tommerup,N. et al. (1993) Hum. Mol. Genet. 2 (10), 1571-1575)
34	7493525CD1	g1017722	4.10E-227	[Homo sapiens] repressor transcriptional factor
35	7021892CD1	g3417312	3.40E-88	[Homo sapiens] RFPL1L (Seroussi,E. et al. (1999) Genome Res. 9 (9), 803-814)

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7492673CD1	316	S100 S124 S247 T37 T87 T139 T186 T207 Y105	N260	signal_cleavage: M1-G31	SPSCAN
					Ribosomal protein S5: V129-A265	HMMER_PFAM
					Ribosomal protein S5 proteins BL00585: K126-I177, V215-C251	BLIMPS_BLOCKS
					Ribosomal protein S5 signature: V129-W192	PROFILESCAN
					RIBOSOMAL PROTEIN S5 30S S2 40S S4 REPEAT S5P LLREP3 PD001364: V129-A265 PD001336: Q79-E128	BLAST_PRODOM
					RIBOSOMAL PROTEIN S5 DM00432 |P25444|92-261: E115-T285 |P15880|92-261: E115-T285 |P27952|92-261: E115-T285 |P49154|78-247: E115-T285	BLAST_DOMO
2	7990930CD1	192	S110 S119 S182 T19 T166 T174 Y180	N7 N108	Ribosomal protein L6: I12-D191	HMMER_PFAM
					Ribosomal protein L6 proteins BL00700: Q8-L45, L62-P100, V112-K121, K141-K184	BLIMPS_BLOCKS
					Ribosomal protein L6 signatures: L55-K121	PROFILESCAN
					ESCHERICHIA COLI RIBOSOMAL PROTEIN L6 DM00422 |P32969|2-185: K2-T186 |P50882|2-183: K2-T186 |P49209|2-189: K2-K184 |Q10232|3-183: V4-G185	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3	7037554CD1	735	S5 S35 S45 S49 S86 S91 S103 S152 S164 S169 S170 S181 S190 S273 S279 S284 S289 S318 S326 S336 S393 S402 S406 S435 S524 S545 S585 S596 S680 T57 T198 T266 T345 Y733	N196 N367 N594	SPLICING YT521-B BRAIN FACTOR PD129722: P493-G681 PD105763: M1-E193 PD129718: R255-A319	BLAST_PRODOM
					DELTA; ENHANCER; CRYSTALLIN; DM07281|P37275|593-734: E158-S273	BLAST_DOMO
					HISTIDINE; SARCOPLASMIC; RETICULUM; CALCIUM; DM07013|P23327|406-624: T93-D282	BLAST_DOMO
					TRICHOHYALIN DM03839|Q07283|91-443: D4-E276, E653-R735	BLAST_DOMO
					EUKARYOTIC; RNA; RNP-1; DM07068|P09406|303-470: R683-R734, P651-R730, E230-R305	BLAST_DOMO
4	1515347CD1	1340	S10 S31 S42 S84 S213 S255 S277 S307 S322 S334 S503 S604 S650 S703 T243 T246 T347 T362 T390 T411 T802 Y206	N134 N221 N576 N641	Helicase conserved C-terminal domain: D115-G198	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4					PROTEIN REPEAT SIGNAL PRECURSOR PRION GLYCOPROTEIN NUCLEAR GPI ANCHOR BRAIN MAJOR PD001091: K721-P788, P844-P1031, P24-P39	BLAST_PRODOM
					ATP NP_BIND DM00266 |P53115|1005-1418: L72-I205 |P22082|788-1207: S84-I205 |P32597|491-911: S84-I205 |A56533|147-551: F82-I205	BLAST_DOMO
					Leucine zipper pattern: L556-L577	MOTIFS
					ATP/GTP-binding site motif A (P-loop): G925-T932	MOTIFS
5	3464492CD1	560	S42 S250 S258 S276 S293 S353 S375 S410 S510 S520 T14 T104 T171 T224 T263 T330 T408 T455 T459 T555 Y115 Y199 Y282 Y302	N33 N234	Helicase conserved C-terminal domain: G82-G168	HMMER_PFAM
					HELICASE ATP-BINDING NUCLEAR RNA MRNA U5 PROCESSING DNA SPLICING SNRNP-SPECIFIC PD099891: L15-N134	BLAST_PRODOM
					SKI2W; SKI2; NUCLEOLAR; HELICASE; DM01537 |P32639|502-912: K10-N197 |P51979|75-389: L23-N197	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6	1794336CD1	436	S281 S393 T238 T322		Zinc finger, C2H2 type: N119-H140, Y398-H417, S65-H87, Y174-H196, Y230-H252, Y314-H336, Y146-H168, Y370-H392, Y202-H224, H342-H364, Y286-H308, Y258-H280	HMMER_PFAM
					C2H2-type zinc finger signature PR00048: P173-R186, L329-G338	BLIMPS_PRINTS
					ssDNA binding protein PF00747: L38-T55, C232-K256, C274-R324, S355-L408	BLIMPS_PFAM
					PATERNALLY EXPRESSED ZN-FINGER PW1 PD017719: G142-F379, G170-F404, G198-H406, C121-K368, C94-I363, C67-D297, E66-F239, G226-E420	BLAST_PRODOM
					NUCLEAR ZINC FINGER TRANSCRIPTION REGULATION REPEAT PD000072: K172-C235, K228-C291, K312-C375, K284-C347, K200-C263, K256-C319, K144-C207	BLAST_PRODOM
					HYPOTHETICAL ZINC FINGER METAL BINDING NUCLEAR PD149420: T162-G338	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6					ZINC FINGER, C2H2 TYPE, DOMAIN DM00002 |Q05481|789-829: Q165-Q206, Q193-E234, R222-E262, R138-E178, Q249-E290, Q277-E318, Q361-C400, Q305-C344 |Q05481|831-885: C151-Q206, C124-E178, C263-E318, V289-R234. C207-E262, C235-E290, C291-P341, C347-P397 |P08042|272-312: R222-E262, Q165-Q206, Q305-C344, R138-E178, Q249-E290, Q333-E374, Q277-E318 |P08042|314-358: C207-H252, C291-H336, C151-H196, C235-H280, C347-H392, C179-H224	BLAST_DOMO
					Zinc finger, C2H2 type, domain: C67-H87, C148-H168, C176-H196, C204-H224, C232-H252, C260-H280, C288-H308, C316-H336, C344-H364, C372-H392	MOTIFS
7	2903694CD1	817	S29 S73 S98 S189 S246 S288 S304 S357 S377 S397 S437 S454 S471 S534 S545 S591 T35 T102 T292 T319 T576 Y41	N14 N282 N712		

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
8	6975426CD1	1273	S128 S274 S399 S451 S496 S550 S557 S590 S599 S600 S641 S644 S755 S832 S911 S968 S977 S982 S1044 S1152 S1173 S1203 T354 T392 T407 T571 T603 T629 T802 T1084 T1111 T1230 T1247 Y129 Y610 Y725 Y883 Y1050 Y1187	N201 N384 N582 N739	Paired amphipathic helix repeat: P141-P187, P322-P381, A477-G523	HMMER_PFAM
					AMPHIPATHIC HELIX TRANSCRIPTION REGULATION REPEAT REPRESSOR SIN3A PD005922: K448-F825 PD133313: M1106-P1273 PD007758: D880-N1107 PD075599: V194-I308	BLAST_PRODOM
					HELIX A REPEAT DM02351 |A56068|216-403: I216-H404 |A56068|42-214: E42-G215	BLAST_DOMO
					SIN3; REGULATORY; DM06689 |P22579|739-1537: I546-A850, L831-E1157, F1169-R1209 |Q09750|444-1412: Q418-P861, L831-T1190	BLAST_DOMO
					Cell attachment sequence: R828-D830	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	4019390CD1	381	S14 S160 S200 T5 T54 T99 T168 T302	N239 N327 N351	KRAB box: L4-K66	HMMER_PFAM
					Zinc finger, C2H2 type: Y229-H251, Y201-H223, Y313-H335, N257-H279, F173-H195, Y341-H363, Y285-H307	HMMER_PFAM
					ZINC FINGER PD01066: F6-G44	BLIMPS_PRODOM
					ZINC FINGER METAL BINDING DNA BINDING PROTEIN ZINC FINGER NUCLEAR TRANSCRIPTION REGULATION REPEAT PD008015: R67-G169 PD001562: L4-W63 PD000072: Y201-C262, K227-C290	BLAST_PRODOM
					PATERNALLY EXPRESSED ZN FINGER PW1 PD017719: D121-F350, G169-H363, A139-D380	BLAST_PRODOM
					KRAB BOX DOMAIN DM00605 |Q03923|1-75: M1-P74 |Q05481|10-83: L4-P74 |P28160|1-69: D8-P74 |S22564|1-63: F13-P74	BLAST_DOMO
					Zinc finger, C2H2 type, domain: C175-H195, C203-H223, C231-H251, C259-H279, C315-H335	MOTIFS
10	986452CD1	290	S3 S17 S142 S280 T58 T76 Y111	N74 N90 N278	Signal_cleavage: M1-G46	SPSCAN
					'Cold-shock' DNA-binding domain: K54-Q84, G85-P94	HMMER_PFAM
					'Cold-shock' DNA-binding domain proteins BL00352: G57-I71	BLIMPS_BLOCKS
					'Cold-shock' DNA-binding domain signature: G37-E135	PROFILESCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
10					Cold shock protein signature PR00050: G57-N72, E78-E87	BLIMPS_PRINTS
					TRANSCRIPTION DNA BINDING REGULATION NUCLEAR REPRESSOR YBOX FACTOR PD003149: G139-E290, G93-Q244 PD004557: G93-N176 PD054259: D20-V59 PD149839: L38-G67	BLAST_PRODOM
					COLD-SHOCK' DNA-BINDING DOMAIN DM02820 |P16990|129-235: G95-N202 |P16990|237-323: M203-E290 |JC2022|42-129: N202-G288 |S48055|206-291: Y207-E290	BLAST_DOMO
11	2807579CD1	588	S13 S18 S54 S94 S125 S142 S189 S197 S222 S261 S268 S432 S442 S455 T25 T34 T64 T100 T182	N63 N129 N195 N311 N324 N440 N463	Fork head domain: K78-D155	HMMER_PFAM
					Fork head domain proteins BL00657: K78-G119, K123-T165	BLIMPS_BLOCKS
					Fork head domain signatures and profile: E19-E103	PROFILESCAN
					Fork head domain signature PR00053: K78-I91, M99-R116, W122-V139	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11					PROTEIN TRANSCRIPTION FACTOR NUCLEAR DNA BINDING REGULATION FORK HEAD FORKHEAD DOMAIN PD000425: K78-E160	BLAST_PRODOM
					FORK HEAD DNA-BINDING DOMAIN DM00381 |I49734|8-281: K78-R170, L20-S85 |A47446|44-314: E70-S173 |P55315|58-332: E70-S173, S310-L376, P357-I390 |A47527|64-311: M16-N207	BLAST_DOMO
					Fork head domain signature 1: K78-I91 signature 2: W122-H128	MOTIFS
12	5724273CD1	103	S9 S25 S58 T42 T49 T96	N7 N80	signal_cleavage: M1-G32	SPSCAN
					KRAB box: V48-K97	HMMER_PFAM
					PROTEIN ZINC FINGER ZINC PD01066: F50-D88	BLIMPS_PRODOM
					ZINC FINGER METAL-BINDING DNA-BINDING PROTEIN FINGER ZINC NUCLEAR REPEAT TRANSCRIPTION REGULATION PD001562: V48-L87	BLAST_PRODOM
					KRAB BOX DOMAIN DM00605 |P52738|3-77: G46-L89 |S42077|18-93: V48-L87 |I48208|18-93: V48-L87 |Q05481|10-83: G46-L89	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
13	3614884CD1	593	S28 S52 S88 S131 S160 S170 S241 S314 S463 S567 T9 T18 T72 T84 T107 T125 T145 T176 T353 T398 T493 T566 Y427		KRAB box: V8-E69	HMMER_PFAM
					Zinc finger, C2H2 type: Y343-H365, F315-H337, F539-H561, F399-Y421, H175-H197, E148-H169, Y203-H225, H259-H281, F483-H505, Y371-H393, F287-H309, Y427-H449, Y455-H477, F231-H253, Y511-H533	HMMER_PFAM
					PROTEIN ZINC FINGER ZINC PD01066: F10-G48	BLIMPS_PRODOM
					ZINC FINGER PROTEIN ZINC FINGER METAL-BINDING DNA-BINDING PD075863: T70-S174	BLAST_PRODOM
					PROTEIN ZINC FINGER METAL-BINDING DNA-BINDING ZINC FINGER PATERNALLY EXPRESSED ZN-FINGER PW1 PD017719: C317-H561, G227-F464	BLAST_PRODOM
					ZINC FINGER DNA-BINDING PROTEIN METAL-BINDING NUCLEAR ZINC FINGER TRANSCRIPTION REGULATION REPEAT PD000072: K201-C264	BLAST_PRODOM
					ZINC FINGER, C2H2 TYPE, DOMAIN DM00002 |Q05481|789-829: K474-E515 |Q05481|831-885: C460-E515 |P08042|314-358: C208-H253	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
13					KRAB BOX DOMAIN DM00605|I48689|11-85: K5-R74	BLAST_DOMO
					Zinc finger, C2H2 type, domain: C205-H225, C233-H253, C261-H281, C289-H309, C317-H337, C345-H365, C373-H393, C429-H449, C457-H477, C485-H505, C513-H533	MOTIFS
14	3794954CD1	281	S3 S34 S53 S75 S190 S269 T98 T162		Zinc finger, C2H2 type: H107-H129, F243-H265, Y135-H157, S163-H185, H209-H237	HMMER_PFAM
					ZINC FINGER DNA-BINDING PROTEIN METAL-BINDING NUCLEAR ZINC FINGER TRANSCRIPTION REGULATION REPEAT PD000072: R105-C168	BLAST_PRODOM
					Zinc finger, C2H2 type, domain: C109-H129, C137-H157, C165-H185, C211-H233, C245-H265	MOTIFS
15	7399016CD1	539	S12 S15 S18 S20 S32 S172 S237 S271 S285 S291 S303 S317 S463 T102 T272 T429 T475		signal_cleavage: M1-A66	SPSCAN
					Zinc finger, C2H2 type: Y421-H443, F479-H502, Y359-H383, L449-H471, R390-H415	HMMER_PFAM
					C2H2-type zinc finger signature PR00048: N420-R433, L436-G445	BLIMPS_PRINTS
					Cytochrome c family heme-binding site signature: C4-G9, C61-A66	MOTIFS
					Zinc finger, C2H2 type, domain: C361-H383, C392-H415, C423-H443, C451-H471, C481-H502	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
16	6996690CD1	390	S91 S134 S264 T114 T249 T348		Zinc finger, C2H2 type: Y312-H334, Y144-H166, Y340-H362, Y284-H306, Y172-H194, Y200-H222, Y368-H390, F256-H278, Y228-H250	HMMER_PFAM
					PROTEIN ZINC FINGER METAL-BINDING DNA-BINDING ZINC FINGER PATERNALLY EXPRESSED ZN-FINGER PW1 PD017719: G84-H334	BLAST_PRODOM
					ZINC FINGER DNA-BINDING PROTEIN METAL-BINDING NUCLEAR ZINC FINGER TRANSCRIPTION REGULATION REPEAT PD000072: K282-C345	BLAST_PRODOM
					MYELOBLAST KIAA0211 ZINC FINGER METAL-BINDING DNA-BINDING PD149061: E145-H334	BLAST_PRODOM
					ZINC FINGER PROTEIN B03B8.4 IN CHROMOSOME III ZINC FINGER DNA BINDING METAL BINDING NUCLEAR PD149420: E141-G308	BLAST_PRODOM
					ZINC FINGER, C2H2 TYPE, DOMAIN DM00002 |Q05481|789-829: R192-E232 |P08042|314-358: C317-H362 |Q05481|831-885: C261-D316 |P52743|31-93: L187-H250	BLAST_DOMO
					Zinc finger, C2H2 type, domain: C146-H166, C174-H194, C202-H222, C230-H250, C258-H278, C286-H306, C314-H334, C342-H362, C370-H390	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
17	7740866CD1	807	S6 S16 S21 S38 S93 S98 S333 S358 S440 S470 S501 S550 S632 S662 S693 S743 T178 T754 T797 Y376 Y568 Y761	N124 N137 N165 N193 N414 N606	Zinc finger, C2H2 type: Y127-H149, Y460-H482, Y404-H426 Y596-H618, Y680-H702, Y155-H177, Y239-H261, C432-H454, Y267-H289, Y323-H345, L295-H317, Y516-H538, Y708-H730, Y568-H590, Y761-H783, Y376-H398, Y652-H674, Y488-H510, Y211-H233, Y624-H646, Y183-H205	HMMER_PFAM
					C2H2-type zinc finger signature PR00048: P459-R472, L639-G648	BLIMPS_PRINTS
					PROTEIN ZINC FINGER METAL-BINDING DNA-BINDING ZINC FINGER PATERNALLY EXPRESSED ZN-FINGER PW1 PD017719: G151-H398, G428-H674	BLAST_PRODOM
					ZINC FINGER DNA-BINDING PROTEIN METAL-BINDING NUCLEAR ZINC FINGER TRANSCRIPTION REGULATION REPEAT PD000072: K458-C521	BLAST_PRODOM
					HYPOTHETICAL ZINC FINGER PROTEIN B03B8.4 IN CHROMOSOME III ZINC FINGER DNA-BINDING METAL-BINDING NUCLEAR PD149420: K214-A346	BLAST_PRODOM
					ZINC FINGER PROTEIN ZINC FINGER METAL-BINDING DNA-BINDING PUTATIVE REX2 TRANSCRIPTION REGULATION PD033163: E131-K265	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
17					ZINC FINGER, C2H2 TYPE, DOMAIN DM00002 |Q05481|831-885: C629-E684 |Q05481|789-829: R644-E684 |P08042|314-358: C629-H674 |P08042|272-312: R452-E492	BLAST_DOMO
					Zinc finger, C2H2 type, domain: C129-H149, C185-H205, C213-H233, C241-H261, C269-H289, C297-H317, C325-H345, C378-H398, C406-H426, C432-H454, C434-H454, C462-H482, C490-H510, C518-H538, C570-H590, C598-H618, C626-H646, C654-H674, C682-H702, C710-H730, C763-H783	MOTIFS
18	8181605CD1	290	S49 S149 T41 T187 T194 Y48 Y73		Zinc finger, C2H2 type: F212-H234, L156-H178, F184-H206, F99-C121	HMMER_PFAM
					Zinc finger, C2H2 type, domain: C158-H178, C186-H206, C214-H234	MOTIFS
19	8266487CD1	452	S31 S172 S181 S209 S219 S339 S393 S430 T131 T142 T205 T249 T387	N263 N298 N299 N361 N391	signal_cleavage: M1-A62	SPSCAN
					Zinc finger, C2H2 type: F130-H154, Y39-H63, F10-H32, F100-H124, F70-H94, S193-H218, F160-H184	HMMER_PFAM
					CheB methylesterase. PF01339: P38-I51	BLIMPS_PFAM
					Zinc finger, C2H2 type, domain: C12-H32, C41-H63, C72-H94, C102-H124, C132-H154, C162-H184, C195-H218	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
20	5552784CD1	259	S153 S178 S195 S253 T59 T78 T139 T232 Y109	N183	signal_cleavage: M1-A57	SPSCAN
					Zinc finger, C3HC4 type (RING finger): C47-C85	HMMER_PFAM
					Zinc finger, C3HC4 type (RING finger), signature: E43-Q93	PROFILESCAN
					ZINC FINGER PROTEIN DNA-BINDING NUCLEAR BMI1 PROTO-ONCOGENE MEL18 FINGER TRANSCRIPTION REGULATION PD007534: N86-K255	BLAST_PRODOM
					ZINC FINGER, C3HC4 TYPE, DM02300 |JC4296|1-222: E33-Y252 |P35226|3-231: R35-K255 |P35227|3-233: R35-K255 |P35820|250-472: V36-E256	BLAST_DOMO
					Zinc finger, C3HC4 type (RING finger), signature: C63-I72	MOTIFS
21	7281230CD1	665	S18 S28 S100 S170 S283 S292 S465 S535 S542 S626 S647 T302 Y301	N210 N315	Zinc finger, C2H2 type: H441-H463, F609-H631, Y637-H659, Y329-H351, F553-H575, F497-H519, Y301-H323, F357-H379, Y385-H407, H413-H435, F469-H491, H525-H547, F581-H603	HMMER_PFAM
					SCAN domain: P27-M122	HMMER_PFAM
					C2H2-type zinc finger signature PR00048: P384-N397, L484-G493	BLIMPS_PRINTS
					PROTEIN ZINC FINGER METAL-BINDING DNA-BINDING ZINC FINGER PATERNALLY EXPRESSED ZN-FINGER PW1 PD017719: G297-F534	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
21					ZINC FINGER METAL-BINDING ZINC FINGER PROTEIN DNA-BINDING NUCLEAR TRANSCRIPTION REGULATION REPEAT PD004640: P27-E149	BLAST_PRODOM
					ZINC FINGER DNA-BINDING PROTEIN METAL-BINDING NUCLEAR ZINC FINGER TRANSCRIPTION REGULATION REPEAT PD000072: K327-C390	BLAST_PRODOM
					ZINC FINGER METAL-BINDING DNA-BINDING PROTEIN ZINC FINGER F18547_1 R28830_2 TRANSCRIPTION REGULATION PD009300: P296-Y385	BLAST_PRODOM
					ZINC FINGER, C2H2 TYPE, DOMAIN DM00002 |P52743|31-93: L316-H379 |P08042|314-358: C362-H407 |Q05481|831-885: C558-P608	BLAST_DOMO
					Eukaryotic thiol (cysteine) proteases cysteine active site: Q34-A45	MOTIFS
					Zinc finger, C2H2 type, domain: C303-H323, C331-H351, C359-H379, C387-H407, C415-H435, C443-H463, C471-H491, C499-H519, C527-H547, C555-H575, C583-H603, C611-H631, C639-H659	MOTIFS
22	7488424CD1	452	S81 S87 S117 S146 S338 S421 T95 T99 T162 T229	N426 N436	SPRY domain: S338-H451	HMMER_PFAM
					B-box zinc finger.: S88-A129	HMMER_PFAM
					Zinc finger, C3HC4 type (RING finger): C15-C55	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
22					RFP TRANSFORMING PROTEIN DM02346: |P19474|59-337: R62-F336 |P15533|63-345: N63-Q264 |A57041|64-348: Q60-G339	BLAST_DOMO
					RFP TRANSFORMING PROTEIN RESPONSIVE FINGER ESTROGEN PHOSPHOPROTEIN DM01944: |P18892|355-477: S338-C448	BLAST_DOMO
					Cell attachment sequence: R303-D305	MOTIFS
					Leucine zipper pattern: L245-L266	MOTIFS
23	7487110CD1	387	S53 S181 S201 S220 S258 T170 T193 T210 T291 Y256	N270 N303	Homeobox domain: R233-R289	HMMER_PFAM
					'Homeobox' domain protein BL00027: L247-R289	BLIMPS_BLOCKS
					'Homeobox' antennapedia-type protein BL00032: R236-T274, Q275-A292	BLIMPS_BLOCKS
					'Homeobox' domain signature and profile: Q246-R309	PROFILESCAN
					Homeobox signature PR00024: Q254-L265, L269-W279, W279-K288	BLIMPS_PRINTS
					HOMEOBOX PROTEIN BH1 HOMEOBOX PROTEIN MBH1 DNA BINDING HOMEOBOX DEVELOPMENTAL PROTEIN NUCLEAR PROTEIN PD173108: L128-P232	BLAST_PRODOM
					HOMEOBOX PROTEIN BH1 HOMEOBOX PROTEIN MBH1 DNA BINDING HOMEOBOX DEVELOPMENTAL PROTEIN NUCLEAR PROTEIN PD171729: T291-R387	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
23					HOMEOBOX PROTEIN BH1 HOMEOBOX PROTEIN MBH1 DNA BINDING HOMEOBOX DEVELOPMENTAL PROTEIN NUCLEAR PROTEIN PD171885: M1-H90	BLAST_PRODOM
					PROTEIN HOMEOBOX DNA BINDING NUCLEAR DEVELOPMENTAL TRANSCRIPTION REGULATION FACTOR HOMEO-DOMAIN METAL BINDING PD000010: R228-Q290	BLAST_PRODOM
					HOMEOBOX DM00009 |A41726|371-433: A229-T291 |P22544|327-389: A229-T291 |B41224|153-215: S224-T291 |P22807|541-604: K231-Q290	BLAST_DOMO
					'Homeobox' domain signature: L265-K288	MOTIFS
24	7495008CD1	255	S15 S77 S133 S151		Helix-loop-helix DNA-binding domain L90-G149	HMMER_PFAM
					Myc-type, 'helix-loop-helix' domain BL00038: E98-R113, G129-G149	BLIMPS_BLOCKS
					Myc-type, 'helix-loop-helix' dimerization domain signature: E114-L170	PROFILESCAN
25	7073515CD1	334	S14 S67 S85 S307 S314	N61 N138	signal_cleavage: M1-S52	SPSCAN
					Homeobox domain: K101-K157	HMMER_PFAM
					'Homeobox' domain protein BL00027: L115-R157	BLIMPS_BLOCKS
					'Homeobox' antennapedia-type protein BL00032: R104-T142, Q143-G160	BLIMPS_BLOCKS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
25					'Homeobox' domain signature and profile: L115-C176	PROFILESCAN
					Homeobox signature PR00024: C122-L133, L137-W147, W147-K156	BLIMPS_PRINTS
					PROTEIN HOMEOBOX DNA BINDING NUCLEAR VENTRAL ANTERIOR HOMEOBOX-CONTAINING PD154879: M1-P100	BLAST_PRODOM
					PROTEIN HOMEOBOX DNA BINDING NUCLEAR VENTRAL ANTERIOR HOMEOBOXCONTAINING PD154490: A265-D334	BLAST_PRODOM
					PROTEIN HOMEOBOX DNA BINDING NUCLEAR VENTRAL ANTERIOR HOMEOBOX-CONTAINING PD154181: D158-L205	BLAST_PRODOM
					PROTEIN HOMEOBOX DNA BINDING NUCLEAR DEVELOPMENTAL TRANSCRIPTION REGULATION FACTOR HOMEO-DOMAIN METAL BINDING PD000010: R99-K157	BLAST_PRODOM
					HOMEOBOX DM00009 |P18488|385-450: R99-Q159 |A46305|134-196: K101-K157 |S60249|130-192: R99-K157 |Q04896|138-201: K101-K157	BLAST_DOMO
					'Homeobox' domain signature: L133-K156	MOTIFS
26	3356640CD1	262	S56 S97 T78 T245 Y194	N148 N152 N260	Zinc finger, C2H2 type: Y194-H216, F110-H132, Y138-H160, Y222-Y244, Y166-R188	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
26					Zinc finger, C2H2 type BL00028: C196-H212	BLIMPS_BLOCKS
					Protein zinc-finger PD00066: H212-C224	BLAST_PRODOM
					ZINC FINGER METAL BINDING DNA BINDING PROTEIN ZINC FINGER NUCLEAR TRANSCRIPTION REGULATION REPEAT PD008015: R3-T105	BLAST_PRODOM
					PROTEIN ZINC FINGER METAL BINDING DNA BINDING ZINC FINGER PATERNALLY EXPRESSED ZN FINGER PW1 PD017719: D58-K251, C112-Y244, K80-F259, N136-F259	BLAST_PRODOM
					ZINC FINGER DNA BINDING PROTEIN METAL BINDING NUCLEAR ZINC FINGER TRANSCRIPTION REGULATION REPEAT PD000072: K164-C227, K192-C255	BLAST_PRODOM
					FINGER; PLACENTAL DM03629 |Q03923|76-132: S12-G69 |C32891|11-67: F18-G69 |Q05481|84-140: S12-N68	BLAST_DOMO
					ZINC FINGER, C2H2 TYPE, DOMAIN DM00002 |Q05481|831-885: C171-Q226, C199-E254, C143-K195, C115-E169	BLAST_DOMO
					Zinc finger, C2H2 type, domain: C112-H132, C196-H216	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
27	2015706CD1	509	S18 S50 S121 S171 S174 S339 S342 S395 S426 S468 S479 T9 T120 T123 T218 T386		Zinc finger, C2H2 type: Y161-H183, Y189-H211, Y217-H239, Y245-H267, Y273-H295, Y301-H323, Y329-H351, Y357-H379, Y385-H407, Y413-H435, H441-H463, Y469-C496	HMMER_PFAM
					KRAB box: L8-T71	HMMER_PFAM
					Zinc finger, C2H2 type BL00028: C331-H347	BLIMPS_BLOCKS
					C2H2-type zinc finger signature PR00048: P384-K397, L456-G465	BLIMPS_PRINTS
					Protein zinc finger PD01066: F10-D48	BLAST_PRODOM
					Protein zinc finger PD00066: H151-C163	BLAST_PRODOM
					PROTEIN ZINC FINGER METAL BINDING DNA BINDING ZINC FINGER PATERNALLY EXPRESSED ZN-FINGER PW1 PD017719: G241-E486	BLAST_PRODOM
					ZINC FINGER DNA BINDING PROTEIN METAL BINDING NUCLEAR ZINC FINGER TRANSCRIPTION REGULATION REPEAT PD000072: K243-C306	BLAST_PRODOM
					ZINC FINGER METAL BINDING DNA BINDING PROTEIN FINGER ZINC NUCLEAR REPEAT TRANSCRIPTION REGULATION PD001562: L8-L47	BLAST_PRODOM
					MYELOBLAST KIAA0211 ZINC FINGER METAL BINDING DNA BINDING PD149061: K190-T396	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
27					ZINC FINGER, C2H2 TYPE, DOMAIN DM00002 |P52743|31-93: L260-H323 |Q05481|789-829: R320-K360 |Q05481|789-829: R320-K360	BLAST_DOMO
					KRAB BOX DOMAIN DM00605|Q05481|10-83: G6-S50 DM00605|Q03923|1-75: G6-S50	BLAST_DOMO
					Zinc finger, C2H2 type, domain: C163-H183, C191-H211, C219-H239, C247-H267, C275-H295, C303-H323, C331-H351, C359-H379, C387-H407, C415-H435, C443-H463	MOTIFS
28	6920755CD1	310	S32 S90 S102 S107 S114 S229 S257 T148 T196 T302		Zinc finger, C2H2 type: F275-H299, F245-H269, H217-H239, I188-H212	HMMER_PFAM
					Zinc finger, C2H2 type BL00028: C219-H235	BLIMPS_BLOCKS
					C2H2-type zinc finger signature PR00048: D216-S229, L262-G271	BLIMPS_PRINTS
					Protein zinc finger PD00066: H235-C247	BLAST_PRODOM
					REX1 PROTEIN REDUCED EXPRESSION1 ZINC FINGER METAL BINDING DNA BINDING TRANSCRIPTION REGULATION REPEAT PD107156: M1-E130	BLAST_PRODOM
					ZINC FINGER METAL BINDING DNA BINDING YY1 TRANSCRIPTION FACTOR TRANSCRIPTIONAL REPRESSOR PROTEIN YIN PD015907: K134-R215	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
28					ZINC FINGER, C2H2 TYPE, DOMAIN DM00002 |P22227|218-246: F236-H265 |A48273|344-372: L237-H265 |P25490|344-372: L237-H265 |P22227|248-280: V266-H299	BLAST_DOMO
					Atp-Gtp-A site: A25-S32	MOTIFS
					Zinc finger, C2H2 type, domain: C190-H212, C219-H239, C247-H269, C277-H299	MOTIFS
29	444179CD1	402	S93 S145 S162 S190 S222 S253 S274 S305 T44 T53 T112 T126 T142 T300	N166 N359	signal_cleavage: M1-L15	SPSCAN
					KRAB box: V43-E105	HMMER_PFAM
					Zinc finger, C2H2 type: Y208-H230, Y292-H314, Y348-H370, Y320-H342, Y376-H398, Y180-H202, H264-H286, N152-H174, Y236-H258	HMMER_PFAM
					Zinc finger, C2H2 type BL00028: C182-H198	BLIMPS_BLOCKS
					Protein zinc finger PD01066: F45-G83	BLAST_PRODOM
					PROTEIN ZINC FINGER METAL BINDING DNA BINDING ZINC FINGER PATERNALLY EXPRESSED ZN-FINGER PW1 PD017719: C154-H398	BLAST_PRODOM
					ZINC FINGER METAL BINDING DNA BINDING PROTEIN FINGER ZINC NUCLEAR REPEAT TRANSCRIPTION REGULATION PD001562: V43-K100	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
29					ZINC FINGER DNA BINDING PROTEIN METAL BINDING NUCLEAR ZINC FINGER TRANSCRIPTION REGULATION REPEAT PD000072: K178-C241	BLAST_PRODOM
					ZINC FINGER PROTEIN 142 KIAA0236 HA4654 TRANSCRIPTION REGULATION DNA BINDING ZINC FINGER METAL BINDING NUCLEAR PD104136: E205-E345	BLAST_PRODOM
					KRAB BOX DOMAIN DM00605 |P52736|1-72: V43-P114 |I48689|11-85: V43-P114 |Q05481|10-83: G41-K100 |Q03923|1-75: G41-R106	BLAST_DOMO
					Zinc finger, C2H2 type, domain: C154-H174, C182-H202, C210-H230, C238-H258, C266-H286, C294-H314, C322-H342, C350-H370, C378-H398	MOTIFS
30	5628380CD1	602	S9 S27 S153 S178 S226 S229 S237 S276 S363 S483 S545 T26 T34 T101 T189 T197 T214 T262 T309 T385 T450 T523 Y445	N116 N173 N441	PROTEIN GRAINY HEAD DNA BINDING ELF1 ELEMENT I BINDING ACTIVITY TRANSCRIPTION FACTOR NTF1 REGULATION NUCLEAR PD144903: F231-A475	BLAST_PRODOM
					TRANSCRIPTION; GLOBIN; CP2; ALPHA DM05518 |P13002|579-1062: F231-A475 |A42030|1-501: E217-R418, V526-K600	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
31	7493789CD1	816	S124 S140 S162 S210 S328 S435 S567 S613 T103 T231 T242 T314 T325 T366 T401 T415 T585 T678 T691 T716 T732 T809	N240 N316 N381	PAZ (proteins Piwi, Argonaut, and Zwille/Pinhead) domain: C192-S328	HMMER_PFAM
					Piwi (a Drosophila protein which functions in RNA interference) domain: L474-V775	HMMER_PFAM
					PROTEIN C ELEGANS ARGONAUTE ZK757.3 CHROMOSOME III INITIATION FACTOR SIMILAR (FACTOR INITIATION BIOSYNTHESIS EIF-2C TRANSLATION PIWI EIF2C CDNA EUKARYOTIC) PD003334: L606-H764	BLAST_PRODOM
					PROTEIN ARGONAUTE EUKARYOTIC INITIATION FACTOR ZWILLE AGO1 LIKE T07D3.7 F48F7.1 TRANSLATION EIF2C PD011593: V384-L599, P541-K612	BLAST_PRODOM
					PROTEIN C ELEGANS C14B1.7 CHROMOSOME III INITIATION FACTOR CODED FOR (FACTOR INITIATION EUKARYOTIC EIF-2C BIOSYNTHESIS EIF2C TRANSLATION ARGONAUTE Q99) PD004358: M170-T383	BLAST_PRODOM
					ARGONAUTE AGO1 LIKE PROTEIN PD128851: V134-G378, V17-P99	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
32	2075194CD1	2248	S28 S42 S57 S73 S86 S100 S138 S144 S153 S197 S199 S221 S340 S379 S397 S426 S431 S432 S505 S568 S597 S626 S652 S664 S674 S678 S682 S700 S770 S809 S827 S874 S912 S955 S994	N136 N377 N539 N847 N864 N1974	Zinc finger, C2H2 type: F1628-H1650, F1714-H1736, F1656-H1680, Y1595-H1618, F1444-H1468, L186-H208, F705-H728, Y1773-H1796, F1567-H1589, F1686-H1708, R850-H873, I1414-H1437, N1537-H1561, F1046-H1070, H215-H238, F557-H581, Y1742-H1767, N1017-H1040, F615-H639	HMMER_PFAM
			S1007 S1076 S1163 S1312 S1329 S1342 S1403 S1407 S1411 S1475 S1636 S1852 S1853 S1881 S1987 S2138 S2214 S2238		Zinc finger, C2H2 type BL00028: V1446-H1462	BLIMPS_BLOCKS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
32			T111 T147 T149 T307 T326 T411 T435 T453 T456 T492 T517 T538 T541 T636 T828 T838 T890 T900 T940 T1041 T1057 T1067 T1118 T1134 T1152 T1212 T1485 T1622 T1629 T1638 T1764 T1826 T1968 T2189 Y1047		C2H2-type zinc finger signature PR00048: E1594-K1607, L1729-G1738	BLIMPS_PRINTS
					HYPOTHETICAL ZINC FINGER PROTEIN B03B8.4 IN CHROMOSOME III ZINC FINGER DNA BINDING METAL BINDING NUCLEAR PD149420: K1440-G1591	BLAST_PRODOM
					Zinc finger, C2H2 type, domain: C188-H208, C246-H268, C617-H639, C1446-H1468, C1488-H1509, C1597-H1618, C1630-H1650, C1658-H1680	MOTIFS
33	2801633CD1	256	S14 S48 S117 S158 T5 T54 T114 T138 T170 T200	N209	KRAB box: L4-K66	HMMER_PFAM
					Zinc finger, C2H2 type Y227-H249, Y199-H221, F171-H193	HMMER_PFAM
					Zinc finger, C2H2 type BL00028: C229-H245	BLIMPS_BLOCKS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
33					C2H2-type zinc finger signature PR00048: P198-S211, L242-G251	BLIMPS_PRINTS
					PROTEIN ZINC FINGER ZINC PD01066: F6-G44	BLIMPS_PRODOM
					ZINC FINGER METAL BINDING DNA BINDING PROTEIN ZINC FINGER NUCLEAR TRANSCRIPTION REGULATION REPEAT PD008015: H68-G167	BLAST_PRODOM
					ZINC FINGER METAL BINDING DNA BINDING PROTEIN FINGER ZINC NUCLEAR REPEAT TRANSCRIPTION REGULATION PD001562: L4-K66	BLAST_PRODOM
					ZINC FINGER DNA BINDING PROTEIN METAL BINDING NUCLEAR ZINC FINGER TRANSCRIPTION REGULATION REPEAT PD000072: K169-C232	BLAST_PRODOM
					KRAB BOX DOMAIN DM00605 |Q05481|10-83: M1-P74 |Q03923|1-75: L4-P74 |P28160|1-69: D8-P74 |S22564|1-63: F13-P74	BLAST_DOMO
					Zinc finger, C2H2 type, domain: C173-H193, C201-H221, C229-H249	MOTIFS
34	7493525CD1	615	S28 S45 S80 S149 S151 S372 S441 S540 T36 T86 T278 T306 T334 T418 T502 T575	N383 N411 N439 N467 N471 N495 N499 N523 N527 N551	KRAB box: L35-K98	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
34					Zinc finger, C2H2 type: Y345-H367, Y289-H311, Y541-H569, F205-H227, Y485-H507, Y457-H479, Y513-H535, Y373-H395, Y233-H255, F261-H283, Y401-H423, Y429-H451, F317-H339	HMMER_PFAM
					Zinc finger, C2H2 type BL00028: C487-H503	BLIMPS_BLOCKS
					PROTEIN ZINC FINGER ZINC PD01066: F37-A75	BLIMPS_PRODOM
					ZINC FINGER METAL BINDING DNA BINDING PROTEIN ZINC FINGER NUCLEAR TRANSCRIPTION REGULATION REPEAT PD008015: R99-G201	BLAST_PRODOM
					PROTEIN ZINC FINGER METAL BINDING DNA BINDING ZINC FINGER PATERNALLY EXPRESSED ZN FINGER PW1 PD017719: G257-F494	BLAST_PRODOM
					ZINC FINGER METAL BINDING DNA BINDING PROTEIN FINGER ZINC NUCLEAR REPEAT TRANSCRIPTION REGULATION PD001562: L35-K98	BLAST_PRODOM
					KRAB BOX DOMAIN DM00605 |Q05481|10-83: L35-P106 |Q03923|1-75: L35-P106 |P28160|1-69: D39-P106 |S22564|1-63: F44-P106	BLAST_DOMO
					Zinc finger, C2H2 type, domain: C207-H227, C235-H255, C263-H283, C291-H311, C319-H339, C347-H367, C375-H395, C403-H423, C431-H451, C459-H479, C487-H507, C515-H535	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
35	7021892CD1	418	S23 S77 S187 S248 S279 S391 T6 T274 T371 Y99	N63 N362	SPRY domain S279-I404	HMMER_PFAM
					Zinc finger, C3HC4 type (RING finger): C142-C183	HMMER_PFAM
					Zinc finger, C3HC4 type BL00518: C157-C165	BLAST_BLOCKS
					RET FINGER PROTEIN-LIKE RFPL1L PD152257: M132-P194	BLAST_PRODOM
					PROTEIN FINGER MIDLINE ZINC FINGER RING STONUSTOXIN PUTATIVE TRANSCRIPTION FACTOR XPRF PD002421: D232-F385	BLAST_PRODOM
					RFP TRANSFORMING PROTEIN DM01944|P14373|368-492: S279-C401 DM01944|P19474|339-465: S279-A386 DM01944|P18892|355-477: S279-F385 DM01944|A43906|483-608: S279-D394	BLAST_DOMO

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
36/7492673CB1/ 1010	1-948, 138-175, 168-205, 640-881, 640-1010
37/7990930CB1/ 612	1-265, 226-552, 403-612
38/7037554CB1/ 2663	1-793, 262-980, 422-901, 441-884, 456-875, 522-761, 567-1207, 669-889, 670-889, 765-1520, 851-1367, 884-1550, 897-1592, 897-1649, 932-1084, 947-1673, 948-1564, 966-1622, 971-1438, 978-1601, 980-1539, 982-1569, 983-1553, 991-1674, 992-1586, 998-1665, 1002-1540, 1003-1547, 1004-1466, 1006-1474, 1006-1674, 1008-1552, 1059-1662, 1061-1558, 1070-1381, 1152-1710, 1156-1591, 1175-1435, 1193-1847, 1195-1749, 1229-1501, 1258-1758, 1279-1797, 1302-1566, 1305-1955, 1468-1957, 1545-2140, 1752-1957, 1766-1999, 1775-1957, 1830-2100, 1890-2317, 1945-2204, 1961-2238, 2025-2203, 2207-2365, 2211-2247, 2211-2250, 2211-2282, 2211-2283, 2211-2302, 2211-2303, 2211-2310, 2211-2319, 2211-2324, 2211-2331, 2211-2334, 2211-2342, 2211-2346, 2211-2352, 2211-2362, 2211-2365, 2211-2369, 2211-2370, 2211-2371, 2211-2387, 2211-2392, 2220-2376, 2250-2539, 2325-2663
39/1515347CB1/ 7188	1-761, 458-1185, 459-1117, 459-1185, 477-870, 477-910, 477-935, 477-943, 477-960, 477-968, 478-976, 485-948, 485-1159, 511-899, 521-971, 529-976, 530-708, 530-918, 532-736, 546-931, 546-946, 561-996, 571-993, 620-1242, 633-1468, 676-1001, 696-868, 907-1407, 943-7188, 2667-3133, 2780-3192, 2787-2913, 2806-3003, 2808-3058, 2808-3217, 2815-2925, 2836-3349, 3058-3593, 3168-3469, 3215-3877, 3296-3918, 3402-3597, 3543-3988, 3598-3622, 3789-4285, 3886-4162, 3886-4218, 3886-4407, 3928-4073, 4101-4560, 4130-4369, 4237-4392, 4247-4623, 4341-4583, 4351-4573, 4351-4869, 4392-4649, 4402-4698, 4416-4960, 4426-4707, 4440-4621, 4478-4975, 4530-5011, 4585-4833, 4633-5216, 4646-4912,
	4662-4869, 4662-4873, 4662-5153, 4735-4991, 4792-5086, 4863-5099, 4913-5223, 4938-5195, 4976-5609, 4994-5247, 4996-5237, 4996-5245, 5000-5277, 5013-5583, 5033-5328, 5039-5283, 5058-5621, 5069-5321, 5083-5271, 5156-5592, 5162-5625, 5169-5632, 5189-5632, 5231-5502, 5231-5506, 5234-5807, 5243-5593, 5243-5630, 5246-5489, 5250-5495, 5250-5615, 5250-5632, 5260-5630, 5269-5532, 5303-5573, 5311-5630, 5313-5632, 5319-5572, 5404-5613, 5404-5632, 5438-5636, 5527-5752, 5611-5846, 5611-5871, 5611-6100, 5700-5942, 5805-6043, 5886-6212, 5969-6184, 5975-6249, 6068-6316, 6146-6372, 6154-6384, 6154-6393

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
40/3464492CB1/ 1972	1-222, 1-572, 143-421, 314-628, 346-873, 413-643, 438-868, 440-760, 514-1330, 555-682, 565-825, 607-1336, 616-1439, 665-930, 682-1112, 682-1178, 694-753, 756-1050, 785-1060, 835-1326, 849-1081, 854-1335, 890-1141, 941-1190, 941-1215, 941-1471, 991-1269, 1027-1312, 1041-1145, 1156-1429, 1177-1972, 1232-1970, 1260-1513, 1314-1611, 1314-1712, 1373-1952, 1507-1971, 1509-1938, 1510-1971, 1526-1945, 1530-1946, 1546-1970, 1550-1945
41/1794336CB1/ 1857	1-1266, 168-972, 202-868, 324-378, 324-551, 324-593, 324-606, 324-611, 324-786, 334-756, 334-792, 334-803, 334-884, 334-885, 334-971, 334-986, 334-1023, 334-1032, 334-1055, 334-1110, 334-1115, 335-551, 335-717, 335-719, 335-768, 335-779, 335-803, 335-885, 340-936, 340-938, 342-611, 346-1148, 356-466, 356-611, 356-627, 356-633, 356-634, 356-803, 356-809, 356-826, 356-853, 356-858, 356-863, 356-885, 357-551, 363-1209, 366-1074, 400-1199, 404-462, 404-466, 404-522, 404-527, 404-971, 405-527, 405-553, 405-611, 405-629, 405-634, 405-881, 406-466, 411-464, 411-801, 412-611, 414-540, 414-543,
	414-593, 414-600, 414-611, 414-629, 414-634, 414-719, 414-769, 414-774, 414-786, 415-527, 415-885, 415-1056, 417-629, 418-714, 419-1052, 419-1162, 423-527, 430-466, 436-552, 436-711, 436-786, 436-885, 438-634, 441-1032, 448-1149, 450-611, 473-800, 481-611, 481-1105, 488-786, 489-611, 489-634, 489-1199, 490-634, 490-858, 493-548, 493-969, 498-634, 498-1200, 500-786, 505-795, 505-971, 507-611, 507-1104, 507-1115, 509-594, 512-881, 513-863, 521-971, 521-1139, 522-761, 531-1200, 534-634, 542-634,
	552-1238, 565-879, 568-1200, 572-1137, 572-1200, 573-634, 573-786, 573-863, 573-1059, 573-1396, 574-878, 574-885, 580-1022, 581-1097, 581-1115, 584-786, 587-1139, 587-1200, 591-634, 598-634, 598-881, 600-1332, 605-1055, 606-850, 631-966, 649-1199, 652-1137, 670-736, 670-786, 670-863, 670-885, 670-938, 670-1023, 670-1137, 671-1133, 671-1137, 673-1128, 674-786, 675-786, 680-729, 681-963, 681-1032, 684-1199, 688-1137, 689-1199, 690-885, 699-1199, 727-863, 727-885, 727-1200, 740-1032, 741-802, 741-863, 741-885, 741-1200, 741-1205, 742-1110, 742-1115, 748-1137, 749-1194, 750-885, 750-1052, 751-1199, 752-774, 754-1047, 754-1199, 755-1199, 756-1387, 759-863, 765-801, 786-885, 800-1436, 813-1402, 825-1032, 826-885, 826-1032, 826-1115, 826-1136, 826-1137, 834-1032, 834-1199, 834-1205, 835-1205, 837-1115, 838-971, 838-1131, 838-1200, 840-1594, 848-885, 850-902, 851-1200,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
41	856-1205, 857-1138, 858-1055, 868-1469, 870-935, 874-935, 904-1515, 908-1138, 908-1191, 909-1032, 909-1115, 909-1138, 909-1199, 910-1115, 910-1194, 910-1199, 913-1199, 918-963, 919-1115, 919-1138, 919-1199, 919-1205, 922-1137, 923-1138, 924-971, 927-1032, 933-1199, 940-1600, 976-1670, 992-1115, 993-1115, 993-1137, 993-1138, 993-1139, 993-1199, 993-1205, 994-1200, 996-1692, 1000-1138, 1002-1131, 1002-1607, 1006-1186, 1006-1199, 1007-1199, 1011-1115, 1019-1857, 1046-1363, 1062-1199, 1063-1138, 1063-1199, 1069-1181, 1072-1136, 1076-1138, 1076-1199, 1077-1199, 1077-1200, 1078-1137, 1078-1199, 1081-1199, 1084-1138, 1086-1137, 1086-1199, 1088-1200, 1089-1200, 1090-1199, 1090-1205, 1091-1199, 1092-1199, 1095-1533, 1101-1137, 1104-1133, 1110-1138, 1136-1199, 1149-1199, 1162-1191, 1162-1200, 1162-1205, 1196-1256, 1196-1262
42/2903694CB1/ 2454	1-2454, 540-1089, 1759-2158
43/6975426CB1/ 4409	1-507, 1-510, 8-448, 17-505, 17-510, 116-510, 145-779, 202-510, 217-510, 279-708, 297-716, 326-875, 343-628, 356-510, 367-510, 371-510, 566-896, 612-832, 783-1415, 810-1375, 1155-1466, 1155-1588, 1158-1960, 1198-1895, 1309-1979, 1377-2083, 1509-1745, 1525-2117, 1550-2023, 1570-1799, 1570-2231, 1623-2295, 1640-2293, 1684-2293, 1707-2343, 1728-2194, 1754-2122, 1831-2310, 1832-2121, 1901-2293, 1947-2243, 2025-2643, 2030-2464, 2086-2235, 2147-2392, 2249-2451, 2250-2567, 2259-2535, 2291-2574, 2327-2512, 2388-3125, 2416-2642, 2419-3094, 2438-3030, 2473-2861, 2473-2935, 2473-2952, 2473-3018, 2473-3047, 2473-3049, 2473-3051, 2473-3072, 2473-3090, 2473-3115, 2473-3317, 2473-3331, 2475-3048, 2477-2952, 2477-3008, 2477-3043, 2477-3071, 2477-3298, 2518-2939,
	2524-2783, 2568-3318, 2574-3131, 2632-2905, 2694-2949, 2696-3043, 2702-3357, 2705-3228, 2705-3234, 2708-3447, 2710-3452, 2710-3525, 2729-3017, 2729-3264, 2730-3317, 2760-3202, 2760-3258, 2760-3507, 2760-3532, 2760-3557, 2761-3050, 2772-3433, 2804-3582, 2807-3063, 2822-3412, 2832-3271, 2834-3248, 2836-3123, 2860-3231, 2860-3325, 2895-3010, 3023-3952, 3057-3746, 3111-3702, 3118-3724, 3118-3849, 3147-3834, 3157-3642, 3196-3925, 3207-3771, 3213-3982, 3216-3889, 3217-3980, 3218-3939, 3226-3899, 3239-3867, 3239-3981, 3246-3862, 3270-3609, 3283-3937, 3320-3891, 3326-3681, 3336-4026, 3349-4083, 3356-4010, 3357-3885, 3368-4062, 3383-3932, 3393-4084, 3436-3722, 3436-3747, 3436-3754, 3436-3769, 3438-3725, 3438-3738, 3452-4409, 3473-3648, 3473-3653, 3486-3749, 3501-4027, 3547-3781, 3563-3834, 3563-3842, 3568-3980, 3569-4160, 3575-4126, 3595-3777

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
44/4019390CB1/ 1290	1-655, 82-170, 82-178, 82-240, 82-258, 82-304, 112-327, 194-537, 292-947, 425-1157, 482-1136, 566-779, 566-785, 566-950, 566-986, 566-1120, 597-863, 597-870, 597-1035, 597-1067, 624-1290, 634-1120, 649-1034, 655-1207, 658-1207, 659-811, 659-953, 677-698, 681-1007, 681-1230, 707-1290, 734-1115, 734-1285, 734-1287, 739-1201, 770-1290, 817-1288, 854-1287, 901-1199, 901-1237, 901-1285, 910-1287, 939-1285, 978-1285, 995-1282, 995-1288, 1017-1289, 1025-1288, 1062-1288, 1070-1288, 1109-1289, 1161-1290, 1162-1289, 1190-1288, 1235-1284, 1264-1285
45/986452CB1/ 1516	1-279, 58-323, 173-470, 259-533, 269-487, 327-441, 331-603, 367-638, 369-657, 510-785, 510-812, 511-733, 511-778, 511-946, 512-750, 512-799, 513-788, 513-795, 517-784, 518-780, 523-794, 523-799, 523-810, 528-751, 534-782, 564-853, 577-776, 597-732, 597-845, 603-869, 603-1349, 607-868, 620-950, 625-834, 627-931, 627-939, 628-814, 628-931, 629-949, 635-1124, 639-874, 639-875, 651-757, 651-924, 651-946, 663-902, 663-921, 665-926, 670-868, 678-942, 688-963, 691-952, 691-1128, 692-915, 692-1024, 695-951, 698-946, 704-1027, 708-911, 710-976, 711-1000, 712-948, 721-1338, 722-969, 724-999, 724-1012, 736-969, 737-915, 737-996, 737-1012, 739-940, 740-1384, 742-1032, 743-943, 757-1010, 757-1019, 761-998, 761-1063, 763-823, 778-938, 784-1005, 786-1003, 790-1335, 815-943, 821-1121, 826-1037, 833-1071, 853-1384, 879-1135, 886-1116, 888-1384, 890-1384, 891-1378, 903-1384, 908-1087, 911-1148, 915-1179, 918-1381, 919-1350, 922-1184, 922-1381, 926-1182, 927-1284, 927-1382, 928-1377,
	928-1384, 944-1381, 946-1384, 948-1220, 948-1381, 949-1382, 955-1371, 955-1377, 955-1381, 959-1384, 968-1384, 969-1374, 969-1381, 970-1384, 971-1262, 972-1384, 974-1384, 975-1371, 976-1235, 977-1377, 980-1242, 981-1384, 983-1384, 989-1221, 989-1281, 989-1377, 992-1377, 993-1237, 993-1377, 993-1384, 996-1344, 1001-1377, 1002-1374, 1002-1377, 1002-1381, 1002-1383, 1002-1384, 1003-1384, 1004-1246, 1005-1384, 1007-1384, 1009-1384, 1010-1264, 1010-1298, 1011-1384, 1012-1381, 1012-1384, 1014-1384, 1015-1384, 1016-1265, 1018-1384, 1020-1282, 1020-1384, 1023-1280, 1024-1384, 1029-1232, 1032-1283, 1033-1381, 1033-1384, 1036-1384, 1040-1280, 1041-1242, 1041-1275, 1041-1283, 1043-1383, 1053-1367, 1055-1309, 1056-1296, 1058-1271, 1058-1295, 1060-1322, 1061-1370, 1063-1303, 1063-1316, 1063-1384, 1064-1353, 1064-1377, 1066-1384, 1067-1354, 1068-1301, 1068-1324, 1070-1383, 1071-1305, 1072-1327, 1074-1384, 1076-1377, 1077-1384, 1078-1280, 1078-1312, 1079-1351,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
45	1080-1300, 1080-1318, 1080-1342, 1080-1351, 1080-1383, 1080-1384, 1081-1361, 1081-1368, 1082-1361, 1082-1384, 1085-1384, 1087-1384, 1088-1384, 1089-1325, 1090-1327, 1096-1377, 1098-1381, 1098-1382, 1099-1377, 1099-1384, 1100-1278, 1100-1383, 1101-1381, 1101-1384, 1104-1384, 1105-1384, 1106-1375, 1108-1384, 1112-1384, 1113-1324, 1116-1384, 1118-1351, 1120-1366, 1121-1364, 1121-1384, 1125-1376, 1128-1371, 1132-1384, 1133-1384, 1134-1379, 1138-1364, 1142-1384, 1146-1259, 1146-1384, 1147-1272, 1150-1384, 1151-1377, 1154-1381, 1155-1381, 1155-1384, 1163-1335, 1163-1384, 1169-1383, 1170-1381, 1170-1384, 1172-1384, 1174-1384, 1175-1384, 1178-1384, 1180-1377, 1181-1384, 1182-1379, 1184-1362, 1184-1516, 1192-1383, 1196-1384, 1198-1333, 1201-1379, 1202-1384, 1206-1384, 1208-1381, 1208-1384, 1210-1384, 1211-1384, 1213-1384, 1216-1377, 1243-1384, 1255-1381, 1260-1350
46/2807579CB1/ 5123	1-314, 1-404, 5-463, 41-463, 59-677, 131-654, 152-463, 160-488, 188-665, 319-623, 332-554, 332-707, 543-1096, 575-1180, 600-1118, 620-1008, 685-1033, 695-1134, 706-836, 707-1213, 744-1056, 772-938, 772-1315, 786-1294, 810-1067, 822-1071, 825-1112, 832-1089, 838-1012, 853-1151, 887-1127, 965-1297, 977-1294, 979-1313, 1009-1267, 1021-1266, 1024-1262, 1047-1321, 1136-1653, 1177-1768, 1177-1804, 1185-1758, 1206-1732, 1218-1429, 1301-1546, 1301-1716, 1310-1410, 1339-2101, 1361-1577, 1401-1952, 1427-1965, 1508-1762, 1552-1967, 1629-2143, 1684-1959, 1702-1836, 1720-2072, 1789-2063, 1790-2191, 1793-2322, 1827-2191, 1829-2289, 1903-2529, 1933-2437, 1948-2215, 1948-2310, 1948-2328,
	1999-2218, 1999-2328, 2043-2261, 2051-2714, 2057-2305, 2085-2321, 2141-2314, 2142-2815, 2164-2812, 2180-2464, 2199-2826, 2221-2826, 2245-2814, 2285-2583, 2388-2656, 2388-2890, 2409-2847, 2428-2685, 2445-2847, 2467-2682, 2476-2743, 2521-2717, 2522-2666, 2532-2795, 2548-2786, 2559-2847, 2584-3128, 2584-3276, 2614-3111, 2615-3221, 2730-3532, 2761-3526, 2787-3322, 2804-3322, 2842-3346, 2872-3322, 2974-3519, 2976-3242, 3028-3609, 3035-3670, 3041-3297, 3047-3325, 3075-3367, 3077-3314, 3083-3335, 3162-3440, 3174-3440, 3176-3342, 3176-3383, 3176-3413, 3176-3451, 3197-3898, 3208-3845, 3225-3344, 3225-3736, 3244-3526, 3277-3533, 3295-3544, 3295-3782,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
46	3295-3846, 3313-3441, 3332-3447, 3333-3866, 3338-3762, 3341-3502, 3342-3604, 3346-3950, 3355-4017, 3367-3425, 3384-3665, 3401-3640, 3401-3657, 3428-3539, 3428-3697, 3434-3744, 3453-3712, 3460-3782, 3489-3748, 3538-3830, 3562-3884, 3583-3815, 3585-3886, 3586-3888, 3586-3900, 3590-3859, 3590-4085, 3644-3911, 3644-4005, 3682-3951, 3700-3954, 3778-4025, 3778-4042, 3848-4050, 3851-4081, 3851-4092, 3851-4418, 3856-4177, 3876-4142, 3918-4151, 3919-4232, 4015-4241, 4045-4346, 4051-4285, 4075-4317, 4096-4355, 4128-4386, 4246-4465, 4246-4474, 4246-4803, 4276-4518, 4276-4527, 4276-4549, 4302-4584, 4302-4897, 4336-4638, 4350-4583, 4429-4671, 4451-5095, 4458-5077, 4483-4760, 4483-5107, 4486-5050, 4494-5095, 4509-5095, 4516-5098, 4547-4738, 4629-4907, 4716-4971, 4719-4875, 4728-4982, 4802-5042, 4802-5043, 4878-5123, 4891-5056
47/5724273CB1/ 707	1-238, 1-453, 1-707, 10-453, 51-313, 63-250, 84-658, 84-707, 94-453, 108-453, 126-370
48/3614884CB1/ 2170	1-525, 111-649, 116-639, 185-756, 239-835, 240-918, 241-841, 245-575, 397-979, 403-984, 423-1046, 487-1030, 559-814, 652-1132, 745-971, 820-1433, 995-1033, 995-1035, 995-1047, 1091-1139, 1162-1201, 1163-1201, 1225-1794, 1225-1795, 1303-1558, 1488-1773, 1500-2075, 1557-2170, 1679-1727, 1750-1789, 1751-1803, 1915-1970
49/3794954CB1/ 2778	1-257, 1-2340, 41-712, 99-335, 111-712, 116-392, 121-420, 121-612, 129-350, 129-507, 224-625, 225-508, 232-839, 341-868, 361-851, 371-848, 378-851, 382-839, 426-852, 459-845, 473-860, 481-792, 481-860, 532-670, 574-848, 677-1324, 677-1341, 677-1350, 677-1391, 757-1391, 1205-1682, 1235-1752, 1244-1416, 1354-1652, 1372-2107, 1424-1761, 1472-1729, 1629-1776, 1633-1822, 1742-1784, 2066-2097, 2066-2108, 2066-2119, 2079-2778, 2132-2265, 2134-2403, 2161-2422, 2161-2752
50/7399016CB1/ 2478	1-511, 101-349, 101-387, 101-649, 123-709, 258-707, 291-545, 327-877, 361-862, 438-720, 460-492, 501-808, 501-924, 511-749, 557-889, 601-1227, 641-1158, 649-1187, 680-1224, 794-1412, 950-1243, 1004-1412, 1024-1384, 1089-1335, 1153-1781, 1268-1463, 1432-1682, 1449-1655, 1469-1690, 1486-2117, 1494-1717, 1544-1714, 1578-1750, 1578-1840, 1588-1845, 1604-1842, 1624-1836, 1654-1831, 1654-2004, 1674-1894, 1674-1930, 1676-1889, 1711-1966, 1711-2006, 1779-2053, 1788-2089, 1788-2278, 1804-1937, 1860-2103, 1888-2459, 1895-2118, 1896-2065, 1899-2457, 1902-2276, 1903-2454, 1910-2187, 1912-2461, 1930-2454, 1981-2197, 1998-2259, 2043-2293, 2167-2446, 2168-2397, 2200-2420, 2290-2478, 2421-2455

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
51/6996690CB1/ 1947	1-500, 1-728, 9-379, 9-517, 9-621, 9-636, 9-641, 9-710, 77-802, 114-724, 114-726, 114-827, 128-654, 196-630, 235-993, 306-1017, 350-1017, 355-592, 436-711, 453-742, 536-1099, 585-631, 585-636, 585-808, 587-783, 587-800, 587-1021, 591-636, 591-802, 591-810, 591-890, 595-641, 595-808, 598-720, 609-636, 616-776, 616-813, 616-892, 619-892, 674-808, 678-783, 678-933, 678-1021, 679-714, 679-716, 679-725, 679-800, 679-892, 682-720, 682-810, 682-977, 687-863, 687-892, 688-808, 688-1058, 691-892, 692-1142, 694-1058, 695-1058, 713-892, 732-808, 746-1019, 751-799, 755-802, 757-1058, 759-800, 759-802, 759-888, 759-1136, 759-1145, 764-810, 772-1037, 772-1280, 826-1058, 831-1058, 832-1021, 833-892, 841-1045, 842-1144, 847-886, 850-888, 850-1224, 856-892, 862-889, 862-1021, 864-1115, 895-1205,
	915-1058, 939-1220, 939-1230, 939-1310, 940-976, 940-1144, 940-1312, 944-1541, 946-1540, 947-1541, 952-1228, 955-1136, 1007-1058, 1007-1140, 1007-1386, 1007-1387, 1007-1541, 1009-1144, 1011-1058, 1011-1228, 1027-1288, 1027-1310, 1049-1621, 1049-1729, 1071-1560, 1079-1144, 1089-1215, 1089-1244, 1089-1312, 1095-1289, 1095-1295, 1095-1312, 1098-1142, 1098-1436, 1099-1220, 1100-1145, 1102-1228, 1108-1228, 1108-1312, 1114-1289, 1116-1310, 1162-1434, 1167-1310, 1169-1228, 1177-1312, 1183-1228, 1186-1304, 1186-1308, 1186-1312, 1192-1436, 1206-1308, 1206-1310, 1207-1310, 1225-1510, 1247-1312, 1247-1313, 1253-1312, 1259-1310, 1262-1436, 1270-1308, 1281-1563, 1290-1436, 1315-1436, 1351-1562, 1351-1583, 1351-1585, 1351-1854, 1351-1947, 1352-1387, 1367-1436, 1410-1568, 1693-1763
52/7740866CB1/ 3553	1-676, 1-3310, 127-620, 160-405, 160-414, 215-414, 226-414, 507-1946, 672-1274, 672-1280, 672-1349, 672-2006, 782-919, 782-943, 782-1251, 782-1316, 803-1316, 811-831, 811-919, 811-921, 811-1000, 811-1011, 813-1087, 813-1230, 848-1000, 852-1004, 852-1252, 856-1386, 861-1252, 864-1011, 866-982, 866-1000, 866-1252, 866-1335, 866-1400, 877-1088, 887-1558, 891-1001, 894-1241, 933-1335, 936-1051, 936-1054, 936-1106, 936-1230, 936-1335, 948-1397, 950-1106, 950-1235, 950-1316, 950-1335, 973-1316, 1045-1252, 1058-1252, 1058-1281, 1058-1365, 1058-1424, 1058-1500, 1058-1556, 1058-1574, 1058-1597, 1058-2154, 1058-2732, 1078-1236, 1078-1240, 1078-1377, 1078-1432, 1078-1436, 1108-1252,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
52	1119-1229, 1119-1252, 1125-1204, 1125-1409, 1125-1578, 1125-1581, 1125-1699, 1127-1400, 1142-1231, 1142-1403, 1142-1432, 1142-1481, 1143-1253, 1149-1898, 1153-1252, 1157-1630, 1161-1574, 1162-1335, 1177-1231, 1185-1315, 1185-1409, 1185-1432, 1189-1315, 1194-1241, 1194-1252, 1200-1347, 1202-1241, 1211-1335, 1260-1794, 1260-1861, 1265-1347, 1272-1712, 1272-1831, 1272-1832, 1286-1481, 1296-1335, 1296-1403, 1297-1352, 1297-1403, 1297-1481, 1328-1432, 1370-1409, 1370-1432, 1370-1481, 1371-1481, 1388-1409, 1388-1432, 1435-1784, 1435-1874, 1529-2061, 1529-2066,
	1564-2172, 1565-2150, 1582-2150, 1598-1756, 1598-1840, 1598-1859, 1598-1918, 1598-1926, 1599-1999, 1600-1663, 1600-1691, 1600-1726, 1603-1999, 1610-2163, 1610-2206, 1610-2275, 1613-1808, 1613-1898, 1613-1998, 1625-1663, 1625-1664, 1625-1666, 1625-1667, 1625-1691, 1625-1720, 1625-1834, 1625-1835, 1626-1666, 1626-1750, 1626-1751, 1632-1674, 1632-1918, 1634-1990, 1637-1736, 1637-1831, 1637-1895, 1637-1958, 1637-1976, 1637-2134, 1637-2522, 1637-2582, 1637-3439, 1641-1999, 1644-2061, 1656-1835, 1680-1915, 1680-1916, 1680-1943, 1680-2002, 1682-2002, 1701-2150, 1713-2420, 1760-1835, 1760-1999, 1760-2000, 1760-2002, 1761-1834, 1764-1924, 1764-1998, 1767-1918, 1767-2061, 1771-1836, 1792-1821, 1792-1834, 1792-1835, 1792-2002, 1792-2003, 1793-2002, 1794-1840, 1794-1918, 1794-1926, 1799-1820, 1799-1927, 1810-1999, 1822-2370, 1822-2450, 1847-2288, 1847-2407, 1847-2408, 1878-2002, 1890-1999, 1907-2002, 1935-2002, 1939-2002, 1958-2002,
	1960-2002, 1961-2002, 1971-2002, 1978-1998, 2008-2450, 2013-2360, 2104-2655, 2104-2821, 2140-2762, 2141-2729, 2149-2311, 2149-2503, 2166-2729, 2168-2327, 2168-2410, 2172-2302, 2172-2311, 2172-2324, 2172-2332, 2172-2419, 2172-2435, 2172-2474, 2172-2494, 2175-2575, 2179-2575, 2186-2736, 2186-2785, 2186-2852, 2189-2384, 2189-2474, 2189-2574, 2201-2239, 2201-2240, 2201-2243, 2201-2267, 2201-2397, 2201-2409, 2201-2411, 2201-2418, 2202-2242, 2202-2326, 2202-2327, 2209-2250, 2209-2419, 2210-2566, 2211-2242, 2211-2243, 2213-2407, 2213-2435, 2213-2471, 2213-2503, 2213-2536, 2213-2712, 2213-3099, 2213-3159, 2213-3439, 2217-2575, 2232-2418, 2233-2494, 2255-2410, 2256-2491, 2256-2492, 2256-2494, 2256-2503, 2256-2574, 2256-2578, 2258-2586, 2277-2729, 2280-2327, 2289-3070, 2293-2574, 2317-2410, 2336-2411, 2336-2584, 2340-2574, 2347-2412, 2349-2494, 2367-2586, 2368-2418, 2369-2575, 2369-2576, 2369-2579, 2369-2586, 2370-2494, 2375-2419,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
52	2386-2586, 2398-2947, 2398-3047, 2423-2984, 2423-2985, 2423-2989, 2454-2503, 2454-2574, 2454-2579, 2482-2578, 2536-2579, 2536-2586, 2537-2579, 2547-2579, 2554-2574, 2584-3027, 2587-2937, 2684-3214, 2684-3232, 2719-3308, 2720-3306, 2745-3306, 2747-2818, 2747-2819, 2747-2822, 2747-2987, 2747-2993, 2749-2822, 2749-2993, 2751-2818, 2751-2973, 2751-2987, 2751-3067, 2754-3151, 2763-3152, 2765-3319, 2765-3372, 2765-3553, 2768-2922, 2789-3143, 2790-2821, 2790-2822, 2792-2879, 2792-3051, 2796-3152, 2833-2901, 2833-2993, 2833-3067, 2833-3155, 2835-3163, 2854-3306, 2857-2984, 2857-3081, 2857-3097, 2857-3290, 2857-3439, 2866-3402, 2915-2988, 2915-3161, 2946-2987, 2946-3152, 2946-3153, 2946-3156, 2946-3163, 2947-3067, 2975-3463, 3000-3409, 3000-3439, 3000-3463, 3060-3155, 3109-3156, 3111-3155, 3112-3163, 3113-3156, 3113-3161, 3124-3156, 3131-3151, 3161-3439, 3164-3439, 3260-3439, 3296-3439, 3297-3439, 3333-3439, 3347-3439, 3370-3439
53/8181605CB1/ 1760	1-676, 114-679, 181-825, 255-685, 437-1120, 483-568, 483-572, 483-587, 483-974, 483-1021, 483-1093, 488-774, 556-914, 575-1057, 576-1190, 627-1154, 627-1333, 933-1730, 947-1571, 1017-1642, 1062-1307, 1144-1760, 1202-1494, 1206-1298, 1206-1357, 1235-1759, 1387-1760
54/8266487CB1/ 2772	1-748, 605-692, 605-934, 605-1395, 609-1209, 819-1103, 819-1285, 819-1385, 820-1147, 822-1279, 872-1436, 872-1480, 893-1147, 893-1206, 922-1588, 1010-1263, 1060-1320, 1157-1811, 1194-1494, 1203-1901, 1242-1447, 1352-2164, 1373-1818, 1378-1668, 1395-1642, 1407-1899, 1700-2191, 1757-2210, 1799-2345, 1799-2430, 1812-2441, 1813-2375, 1813-2441, 1814-2247, 1825-2337, 1918-2610, 1954-2503, 1954-2504, 2023-2270, 2023-2615, 2023-2686, 2135-2660, 2187-2415, 2251-2669, 2285-2475, 2351-2686, 2459-2671, 2481-2772, 2572-2758
55/5552784CB1/ 1151	1-256, 108-769, 152-1144, 238-941, 252-851, 253-886, 255-720, 255-910, 257-867, 270-832, 271-670, 271-867, 271-870, 273-837, 282-807, 287-643, 287-955, 298-755, 298-784, 299-907, 300-940, 327-1060, 327-1101, 354-708, 360-616, 364-1144, 366-1017, 366-1102, 367-1132, 391-754, 394-908, 394-971, 398-674, 398-1151, 404-696, 413-923, 416-777, 418-1064, 418-1132, 442-687, 442-909, 452-706, 461-743, 461-758, 464-917, 468-1132, 469-1126, 478-1059, 484-1140, 508-778, 519-723, 519-795, 525-724, 526-872, 532-908, 532-1011, 536-777, 536-792, 536-972, 536-978, 536-993, 536-997, 536-1018, 536-1027, 536-1031, 536-1055, 536-1089, 536-1102, 536-1105, 536-1111, 536-1120, 536-1125, 536-1128, 536-1132, 536-1135, 536-1142, 536-1143, 536-1151, 539-1111, 540-781, 540-1151, 545-1069, 547-907, 547-1017, 550-907, 555-1151, 558-765, 561-807, 561-1144, 565-1047, 572-1151, 576-1151, 815-1091

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
56/7281230CB1/ 2230	1-606, 1-618, 158-625, 160-625, 170-407, 622-721, 622-1183, 622-2230, 1056-1531, 1062-1851, 1121-1912, 1121-2002, 1122-1246, 1138-1498, 1139-1246, 1145-2086, 1148-1246, 1184-1750, 1187-1240, 1307-1414, 1350-2164, 1355-1417, 1355-1666, 1396-2187, 1599-1750, 1604-2170, 1605-1918, 1649-2086, 1726-2203, 1773-2086, 1859-2170, 1988-2086, 2111-2164
57/7488424CB1/ 1976	1-525, 1-630, 2-575, 2-646, 2-680, 25-1976, 96-583, 156-713, 188-893, 197-686, 214-844, 237-723, 243-996, 268-910, 269-769, 284-725, 285-793, 288-802, 313-883, 313-965, 334-844, 342-880, 347-861, 358-625, 368-956, 372-958, 404-1089, 412-1026, 418-910, 449-1072, 457-958, 497-956, 553-650, 573-1112, 579-1106, 608-1308, 609-1072, 671-1233, 680-1188, 689-1308, 728-1222, 798-926, 798-1086, 803-1189, 1120-1661, 1376-1679
58/7487110CB1/ 1357	1-1164, 1-1357, 83-968, 664-740, 711-851, 715-851, 852-983, 1106-1357
59/7495008CB1/ 2153	1-387, 1-2145, 12-115, 667-1306, 706-912, 712-912, 736-1261, 736-1266, 742-1001, 742-1334, 745-991, 745-1350, 755-912, 771-1348, 776-1004, 779-1439, 802-1000, 802-1246, 829-912, 830-1057, 842-908, 842-1475, 882-1464, 885-1183, 891-1231, 892-1213, 896-1124, 900-1088, 925-1131, 925-1261, 976-1200, 977-1720, 983-1293, 991-1255, 998-1231, 998-1245, 1012-1242, 1017-1275, 1049-1281, 1050-1318, 1064-1734, 1289-1788, 1328-1959, 1374-1867, 1395-2131, 1437-2094, 1446-2042, 1448-1683, 1459-2005, 1459-2011, 1461-2143, 1504-1794, 1504-1860, 1515-2150, 1542-2135, 1543-1846, 1544-2109, 1554-1810, 1567-1807, 1576-2133, 1606-1808, 1606-2132, 1606-2145, 1617-2138, 1626-1858, 1631-2008, 1660-2145, 1671-2153, 1672-1919, 1681-1928, 1688-2146, 1693-1950, 1697-2151, 1703-1848, 1703-2145, 1703-2153, 1708-2142, 1710-1965, 1713-2145, 1715-2145, 1723-1957, 1724-1970, 1724-2153, 1726-1851, 1726-1943, 1726-1954, 1726-2028, 1726-2146, 1727-1836, 1732-2145, 1756-2152, 1761-2152
60/7073515CB1/ 1104	1-226, 22-753, 607-1002, 869-1104, 870-1104, 871-1104, 890-1005
61/3356640CB1/ 2597	1-265, 1-287, 4-265, 4-555, 9-626, 31-354, 34-282, 38-328, 351-669, 365-642, 475-674, 511-1069, 548-855, 624-757, 624-842, 624-932, 624-979, 637-842, 694-833, 694-856, 694-976, 694-1033, 694-1140, 694-1264, 694-1268, 699-758, 699-799, 699-842, 699-846, 699-899, 699-938, 699-945, 699-979, 699-1064, 699-1068, 699-1108, 699-1181, 699-1184, 699-1201, 699-1346, 699-1352, 703-762, 703-805, 703-812, 703-815, 703-851, 703-865, 703-889, 703-929, 703-932, 703-979, 703-1097, 703-1100, 703-1135, 703-1196, 703-1291, 703-1302, 707-984, 707-1108, 707-1195, 708-781, 708-793, 708-814, 708-833, 708-841,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
61	708-842, 708-848, 708-883, 708-884, 708-896, 708-916, 708-928, 708-948, 708-974, 708-1013, 708-1014, 708-1027, 708-1046, 708-1064, 708-1066, 708-1095, 708-1102, 708-1150, 708-1184, 708-1219, 708-1221, 708-1263, 708-1268, 708-1277, 709-752, 709-762, 709-771, 709-846, 709-1010, 709-1046, 715-865, 715-867, 715-878, 716-1277, 717-759, 718-928, 733-1271, 738-858, 738-899, 738-903, 738-905, 738-916, 738-930, 738-1009, 738-1108, 738-1277, 740-781, 740-878, 740-889, 740-935, 740-1001, 742-1046, 754-1228, 754-1354, 757-1424, 760-1066, 765-1236, 774-1435, 775-1436, 779-878, 779-1039, 779-1346,
	783-1013, 783-1352, 784-989, 784-1010, 784-1387, 791-916, 791-1105, 791-1360, 795-968, 795-1012, 795-1130, 795-1142, 798-1400, 802-1012, 804-863, 810-878, 811-1007, 812-1001, 813-1013, 817-973, 817-1086, 817-1214, 822-1130, 822-1369, 822-1371, 826-1130, 827-1130, 838-1316, 838-1436, 841-1517, 841-1520, 844-1107, 844-1427, 849-1107, 859-976, 859-1096, 859-1224, 859-1319, 859-1543, 861-1012, 867-1010, 867-1109, 867-1224, 867-1436, 868-1066, 868-1307, 871-1453, 872-1013, 881-1096, 885-1007, 885-1046, 885-1271, 891-1491, 895-1100, 896-1086, 901-1056, 901-1169, 901-1177, 902-1046, 906-1012, 906-1443, 910-1013, 910-1214, 910-1271, 919-1520, 920-989, 922-1400, 922-1529, 926-1234, 926-1538, 928-1601, 928-1602, 933-1046, 933-1180, 933-1212, 933-1325, 934-1182, 934-1382, 935-1601, 941-1180, 942-1396, 946-1103, 946-1319, 951-1192, 952-989, 953-1307, 960-1533, 961-1307, 965-1180,
	969-1130, 972-1010, 972-1529, 980-1182, 981-1130, 982-1046, 982-1261, 983-1253, 990-1529, 990-1536, 992-1130, 994-1307, 1005-1723, 1006-1066, 1006-1478, 1010-1318, 1010-1590, 1014-1265, 1015-1279, 1015-1409, 1016-1279, 1017-1285, 1017-1685, 1025-1265, 1026-1487, 1027-1687, 1031-1066, 1031-1220, 1031-1277, 1031-1601, 1034-1601, 1035-1178, 1035-1276, 1035-1346, 1036-1241, 1039-1600, 1053-1193, 1053-1214, 1058-1108, 1058-1654, 1062-1382, 1063-1214, 1063-1259, 1065-1264, 1066-1225, 1066-1345, 1070-1253, 1075-1382, 1075-1396, 1075-1598, 1076-1214, 1076-1335, 1078-1382, 1090-1150, 1090-1382, 1090-1564, 1090-1687, 1094-1369, 1094-1394, 1096-1768, 1100-1350, 1100-1543, 1102-1214, 1120-1279, 1120-1346, 1120-1360, 1120-1476, 1120-1565, 1120-1687, 1120-1768, 1120-1807, 1123-1262, 1123-1476, 1123-1697, 1128-1380, 1128-1683, 1129-1443, 1129-1466, 1144-1193, 1144-1346, 1148-1307, 1148-1335, 1148-1352, 1149-1350, 1150-1276, 1150-1427, 1150-1738,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
61	1151-1420, 1153-1706, 1154-1697, 1158-1466, 1160-1307, 1162-1256, 1162-1262, 1162-1263, 1162-1271, 1162-1466, 1174-1241, 1174-1652, 1174-1781, 1178-1443, 1178-1486, 1178-1790, 1180-1852, 1185-1420, 1185-1464, 1185-1560, 1198-1360, 1199-1656, 1199-1882, 1202-1768, 1203-1443, 1212-1355, 1212-1551, 1212-1783, 1213-1543, 1215-1271, 1221-1382, 1224-1277, 1224-1808, 1230-1466, 1230-1498, 1231-1382, 1232-1420, 1234-1285, 1234-1513, 1235-1505, 1241-1382, 1244-1325, 1244-1543, 1244-1788, 1246-1543, 1257-1975, 1258-1543, 1258-1734, 1258-1852, 1262-1536, 1262-1566, 1266-1937, 1266-1940, 1268-1518, 1268-1718, 1270-1516, 1270-1535, 1270-1732, 1277-1516, 1291-1436, 1291-1472, 1291-1528, 1291-1656, 1291-1867, 1296-1536, 1296-1851, 1296-1852, 1298-1543, 1305-1466, 1306-1516, 1308-1367, 1308-1906, 1314-1382, 1315-1520, 1317-1518, 1321-1477, 1321-1589, 1321-1597, 1322-1466, 1326-1629, 1326-1842, 1326-1871, 1330-1629, 1331-1400, 1342-1818,
	1342-1950, 1345-1995, 1345-1997, 1365-1555, 1365-1600, 1365-1718, 1366-1466, 1366-1627, 1366-1654, 1366-1935, 1367-1814, 1371-1601, 1371-1940, 1380-1523, 1380-1564, 1380-1685, 1380-1949, 1381-1718, 1386-2005, 1387-1543, 1389-2022, 1390-1600, 1392-1443, 1392-1523, 1401-1561, 1402-1588, 1402-1601, 1402-1681, 1403-1674, 1404-1802, 1405-1589, 1409-1543, 1412-1957, 1414-1523, 1414-1718, 1428-1988, 1429-1988, 1429-2026, 1432-1687, 1432-1904, 1432-1985, 1437-1687, 1437-1908, 1438-1802, 1446-1684, 1446-1714, 1446-1812, 1446-1984, 1449-1706, 1449-1718, 1456-1693, 1456-2022, 1464-1686, 1464-1768, 1464-1882, 1464-1984, 1482-1535, 1482-1536, 1482-2030, 1483-1590, 1483-1629, 1483-1687, 1484-1674, 1489-1598, 1489-1757, 1489-1765, 1490-1949, 1494-1812, 1494-2021, 1496-1629, 1498-1802, 1510-1980, 1510-1988, 1514-1816, 1514-2022, 1516-2179, 1520-1766, 1520-1970, 1522-1779, 1522-1984, 1539-1988, 1548-1682, 1548-1768, 1548-1787, 1548-1851,
	1548-1852, 1548-2095, 1558-1768, 1558-1785, 1568-1768, 1569-1718, 1569-1763, 1570-1634, 1570-1729, 1570-1849, 1570-1980, 1571-1841, 1579-1946, 1580-1718, 1594-2026, 1594-2210, 1598-1906, 1598-2179, 1602-1858, 1603-1889, 1603-2251, 1604-1869, 1605-1840, 1606-1970, 1611-1857, 1615-2208, 1615-2215, 1617-1808, 1617-2195, 1624-1852, 1624-2179, 1624-2215, 1632-1766, 1632-1812, 1632-1850, 1632-1987, 1632-1990, 1632-2021, 1632-2194, 1633-1970, 1635-2022, 1642-1788, 1642-1802, 1644-1703, 1644-2202, 1651-1840, 1652-1841, 1653-1852, 1654-1780, 1654-1933, 1655-1925, 1663-2202, 1663-2204, 1664-1802, 1666-1970, 1678-2149, 1678-2211, 1682-1957, 1682-1990, 1684-2215, 1689-1938, 1689-1946, 1701-1891, 1701-1936, 1701-1973, 1701-2018, 1701-2215, 1702-1947, 1702-2215, 1703-2153, 1703-2208, 1707-1948, 1707-2215, 1711-2215, 1726-1936, 1728-1787, 1728-2211, 1736-1812, 1736-1925, 1736-1940, 1737-1938, 1738-1850, 1738-2017, 1739-1999, 1746-1808,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
61	1746-1937, 1746-2030, 1747-1812, 1750-2021, 1762-1812, 1762-2201, 1762-2215, 1772-2023, 1772-2223, 1772-2321, 1792-2030, 1792-2215, 1792-2223, 1800-1976, 1800-1980, 1804-2521, 1805-1980, 1824-1970, 1824-1984, 1824-2022, 1824-2091, 1824-2215, 1825-1990, 1826-1970, 1830-1943, 1830-1980, 1830-1988, 1830-2182, 1830-2215, 1857-2030, 1858-2123, 1858-2151, 1858-2215, 1858-2223, 1868-1980, 1877-2223, 1880-2502, 1885-1970, 1885-1997, 1885-1998, 1885-2018, 1885-2190, 1885-2208, 1885-2215, 1887-1943, 1889-1943, 1890-2215, 1891-1997, 1891-2021, 1893-2215, 1894-1997, 1894-2018, 1896-1946, 1896-1955, 1896-1988, 1902-1947, 1903-2085, 1905-1990, 1906-2026, 1906-2178, 1907-1946, 1907-1997, 1907-2170, 1914-1990, 1914-2215, 1918-1970, 1918-2423, 1930-2440, 1934-1984, 1934-2204, 1934-2215, 1937-2574, 1940-1980, 1942-1997, 1970-2572, 1970-2586, 1995-2536,
	2024-2527, 2052-2151, 2052-2179, 2052-2182, 2052-2215, 2053-2208, 2054-2091, 2054-2098, 2054-2153, 2054-2176, 2054-2193, 2054-2208, 2054-2215, 2054-2432, 2055-2193, 2055-2208, 2057-2098, 2064-2210, 2064-2215, 2065-2215, 2065-2432, 2066-2215, 2067-2123, 2067-2210, 2067-2215, 2068-2208, 2069-2215, 2074-2475, 2077-2215, 2097-2587, 2101-2215, 2102-2215, 2102-2223, 2102-2432, 2103-2215, 2109-2215, 2115-2215, 2119-2215, 2129-2214, 2130-2215, 2138-2182, 2138-2196, 2141-2215, 2147-2215, 2148-2223, 2150-2215, 2151-2208, 2151-2223, 2152-2215, 2163-2215, 2175-2215, 2190-2535, 2195-2241, 2195-2250, 2226-2417, 2226-2597, 2250-2586, 2278-2544, 2278-2575, 2278-2597, 2285-2426, 2288-2423, 2288-2427, 2289-2356, 2295-2501, 2298-2432, 2301-2351, 2301-2387, 2301-2426, 2301-2427, 2301-2432, 2304-2426, 2312-2345, 2323-2351, 2323-2389, 2323-2399, 2323-2426, 2323-2432, 2333-2590, 2335-2426, 2376-2427, 2376-2432
62/2015706CB1/ 1959	1-457, 47-484, 137-792, 200-452, 284-484, 544-1322, 545-817, 554-812, 554-1020, 599-1269, 599-1302, 789-816, 789-831, 789-937, 789-1163, 789-1182, 791-900, 791-935, 791-1049, 791-1076, 792-833, 792-1100, 792-1103, 798-855, 798-915, 798-1121, 798-1123, 799-841, 799-917, 799-1076, 799-1367, 803-863, 803-905, 803-915, 803-929, 803-1181, 805-841, 805-1019, 805-1182, 862-937, 862-1105, 866-1076, 876-915, 876-1187, 887-929, 887-1098, 887-1181, 888-917, 888-1014, 888-1098, 888-1167, 888-1433, 889-917, 942-1076, 943-1076, 949-1404, 966-1103, 966-1182, 966-1522, 979-1076, 979-1182, 979-1433, 990-1076, 991-1076, 997-1068, 997-1169, 997-1419, 999-1419, 1000-1518, 1001-1527, 1014-1235, 1014-1443, 1036-1619, 1045-1076, 1045-1108, 1045-1123, 1045-1181, 1050-1266, 1051-1328, 1056-1096, 1056-1182, 1056-1266, 1057-1522, 1066-1181, 1081-1181, 1110-1903, 1129-1167, 1129-1182, 1129-1253, 1129-1266, 1129-1359, 1129-1433, 1129-1606, 1141-1167, 1141-1177, 1141-1182, 1141-1273,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
62	1141-1409, 1141-1518, 1141-1527, 1147-1527, 1221-1419, 1221-1421, 1221-1439, 1221-1527, 1221-1693, 1223-1433, 1223-1575, 1223-1601, 1224-1335, 1224-1498, 1224-1587, 1224-1601, 1225-1266, 1249-1527, 1249-1656, 1251-1751, 1252-1671, 1253-1902, 1266-1434, 1266-1649, 1287-1335, 1287-1359, 1287-1433, 1287-1522, 1287-1693, 1288-1779, 1291-1433, 1294-1553, 1294-1932, 1294-1940, 1294-1958, 1295-1439, 1295-1580, 1302-1518, 1302-1601, 1305-1488, 1305-1505, 1305-1606, 1366-1522, 1370-1669, 1371-1505, 1371-1777, 1377-1497, 1377-1770, 1391-1433, 1391-1606, 1391-1777, 1392-1518, 1393-1434, 1393-1779, 1431-1687, 1447-1720, 1455-1601, 1459-1959, 1464-1601, 1465-1522, 1465-1527, 1465-1601, 1465-1693, 1470-1589, 1470-1795, 1475-1606, 1477-1518, 1483-1580, 1486-1782, 1495-1804, 1503-1671, 1504-1682, 1530-1580, 1543-1580, 1549-1606, 1549-1777, 1557-1601, 1559-1601, 1559-1661, 1559-1671, 1559-1779, 1624-1779, 1715-1959, 1717-1755, 1717-1772
63/6920755CB1/ 1401	1-1059, 51-1401, 127-1059, 222-704, 222-731, 313-781, 487-731
64/444179CB1/ 3406	1-684, 1-715, 404-586, 404-831, 488-1287, 598-1287, 629-1435, 667-1388, 713-3406, 723-1287, 741-987, 741-1236, 897-1025, 898-1028, 898-1107, 909-1191, 912-1219, 912-1252, 912-1330, 912-1446, 914-1342, 923-1010, 923-1037, 923-1113, 960-1028, 960-1107, 975-1025, 976-1191, 981-1015, 981-1115, 982-1028, 982-1036, 982-1056, 982-1077, 982-1191, 987-1191, 997-1191, 1044-1112, 1044-1191, 1077-1113, 1078-1191, 1149-1183, 1149-1454, 1150-1191, 1153-1968, 1155-1191, 1159-1479, 1161-1437, 1161-1443, 1165-1191, 1168-1437, 1168-1443, 1175-1443, 1186-1819, 1196-1618, 1233-1373, 1233-1457, 1233-1538, 1233-1653, 1235-1795, 1238-1819, 1259-1449, 1259-1450, 1259-1456, 1259-1485, 1259-1495, 1323-1372, 1323-1388, 1367-1968, 1368-1625, 1400-1495, 1401-1529, 1402-1604, 1402-1611, 1417-2040, 1417-2061, 1427-1484, 1427-1496, 1427-1508, 1427-1647, 1427-1663, 1479-1529, 1564-1663,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
64	1566-2186, 1570-1665, 1573-1617, 1574-1663, 1575-1624, 1575-1640, 1581-1617, 1585-1663, 1675-1964, 1713-2188, 1781-2307, 1781-2434, 1803-2340, 1885-2325, 1885-2602, 1900-2421, 1995-2507, 2018-2232, 2048-2307, 2049-2316, 2138-2278, 2138-2610, 2199-2454, 2227-2811, 2232-2471, 2268-2825, 2282-2618, 2290-2730, 2297-2834, 2346-2860, 2347-2840, 2366-2629, 2435-2681, 2452-2697, 2454-2663, 2473-2763, 2522-3225, 2574-3213, 2586-3076, 2613-3212, 2620-3215, 2620-3225, 2620-3235, 2635-2896, 2644-2920, 2651-3243, 2664-3225, 2671-2938, 2693-3254, 2695-3213, 2706-3235, 2755-3008, 2755-3323, 2777-3247, 2782-3253, 2784-3254, 2788-3406, 2794-3246, 2799-3406, 2802-3246, 2804-3246, 2823-3060, 2824-3259, 2825-3248, 2831-3247, 2845-3246, 2845-3248, 2847-3250, 2853-3247, 2855-3246, 2856-3128, 2858-3106, 2887-3180, 2887-3181, 2888-3087, 2962-3248, 2981-3249, 2981-3255, 2981-3320, 2981-3331, 3045-3245
65/5628380CB1/ 2718	1-621, 198-633, 206-599, 312-545, 580-877, 580-909, 652-886, 652-946, 652-1144, 677-927, 721-935, 740-1379, 1119-1396, 1119-1672, 1135-1765, 1162-1831, 1215-1879, 1282-1485, 1340-1857, 1340-1858, 1361-1963, 1389-1676, 1458-2020, 1498-1998, 1524-2167, 1607-2155, 1641-1979, 1680-1948, 1724-2155, 1748-2277, 1823-2082, 1837-2519, 1853-2500, 1883-2508, 1964-2181, 1964-2470, 1979-2678, 2011-2658, 2018-2219, 2023-2706, 2054-2624, 2059-2564, 2059-2718, 2060-2296, 2066-2593, 2104-2541, 2117-2678, 2154-2662, 2165-2685, 2174-2437, 2175-2516, 2177-2395, 2180-2714, 2187-2500, 2192-2673, 2248-2699, 2261-2700, 2263-2718, 2270-2690, 2357-2718, 2418-2691, 2565-2698
66/7493789CB1/ 3325	1-729, 152-786, 231-796, 252-886, 260-715, 270-881, 322-781, 413-801, 485-801, 600-1138, 887-1157, 887-1226, 1085-1719, 1441-1687, 1441-1927, 1496-1762, 1506-1905, 1781-2320, 2035-2568, 2043-2310, 2043-2543, 2094-2640, 2174-2405, 2174-2800, 2207-2442, 2222-2794, 2224-2533, 2354-2954, 2545-3001, 2590-3289, 2629-3117, 2657-2956, 2683-3133, 2717-3325
67/2075194CB1/ 8114	1-788, 54-654, 54-762, 57-639, 406-948, 704-1006, 704-1007, 875-1117, 905-1339, 940-1339, 969-1341, 1044-1336, 1174-1609, 1174-1932, 1219-1289, 1220-4938, 1349-1822, 1390-2040, 1394-4823, 1620-2317, 1906-2495, 1906-2501, 1906-2572, 2038-4823, 2238-2808, 2695-3220, 2695-3278, 2758-4823, 2802-3050, 2858-4823, 2958-4823, 3008-4823, 3058-3765, 3058-4823, 3144-3332, 3149-3765, 3180-3765, 3383-8083, 5011-5386, 5011-5564, 5350-5863, 5350-6031, 5563-7943, 5587-5752, 5694-6326, 5773-6330, 5774-6330, 5779-6330, 5821-6330, 5822-6330, 5869-6744, 6050-6328, 6050-6425, 6069-6330, 6150-6834, 6205-6595, 6208-6861, 6338-6761, 6378-6634, 6441-6689, 6441-6864, 6622-7372, 6683-7251, 6824-7075, 6854-7155, 6868-7191, 6868-7203, 6868-7387, 6928-7348, 6992-7499, 7037-7661,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
67	7044-7665, 7114-7748, 7156-7512, 7188-7522, 7206-7744, 7225-7743, 7235-7787, 7237-7840, 7365-7620, 7365-7892, 7382-7571, 7394-7896, 7396-8074, 7399-8098, 7407-7687, 7417-7836, 7427-7908, 7451-8076, 7470-8060, 7496-7972, 7503-8074, 7503-8095, 7520-7799, 7520-7992, 7520-8078, 7525-8075, 7535-8053, 7544-8102, 7608-8114, 7614-8099, 7614-8105, 7643-7842, 7667-8084, 7700-8085, 7706-8085, 7708-7933, 7714-8102, 7721-8057, 7734-8082, 7737-8085, 7783-8078, 7797-8058, 7812-8049, 7823-7874, 7823-7875, 7824-7875, 7858-8064, 7871-8083, 7875-7910, 7875-7924, 7875-7925, 7876-7925, 7950-8105
68/2801633CB1/ 1530	1-250, 1-382, 1-471, 1-485, 1-489, 1-498, 1-552, 1-553, 1-555, 1-556, 1-571, 1-597, 1-609, 1-614, 24-572, 27-663, 29-300, 29-389, 30-305, 37-708, 56-585, 60-770, 65-667, 139-741, 142-800, 144-689, 162-822, 202-700, 217-680, 231-939, 252-893, 257-884, 257-901, 260-915, 262-664, 292-669, 340-613, 399-761, 407-680, 498-1176, 522-1180, 669-800, 669-811, 669-831, 669-849, 669-857, 691-809, 724-766, 724-833, 746-833, 748-833, 753-873, 753-906, 753-941, 775-934, 776-857, 858-1530, 860-945, 892-929, 892-933, 892-945, 892-967, 892-999, 910-1512, 916-1344, 969-1258, 1000-1253
69/7493525CB1/ 2026	1-546, 1-568, 18-296, 75-130, 75-924, 76-675, 76-786, 76-807, 76-832, 76-838, 76-840, 76-914, 76-928, 76-949, 76-953, 80-709, 80-947, 216-947, 274-923, 312-976, 445-1106, 447-812, 447-1096, 453-985, 454-1332, 462-1015, 490-979, 496-979, 525-1305, 555-979, 658-692, 659-694, 700-1124, 700-1302, 704-742, 704-746, 704-756, 704-849, 704-920, 704-937, 704-956, 704-1102, 704-1106, 704-1178, 704-1318, 705-761, 706-753, 707-753, 707-756, 707-934, 707-964, 707-993, 707-1017, 707-1057, 719-753, 720-753, 720-756, 720-853, 720-898, 720-901, 720-912, 720-913, 720-934, 720-937, 720-967, 720-997, 720-1012,
	720-1018, 720-1067, 720-1069, 720-1107, 720-1142, 720-1215, 720-1225, 720-1234, 720-1274, 723-1142, 725-1486, 727-1234, 790-841, 790-889, 790-899, 790-982, 790-983, 790-1012, 790-1066, 790-1107, 790-1108, 790-1184, 790-1194, 790-1266, 790-1309, 791-850, 791-853, 791-877, 791-892, 791-895, 791-901, 791-925, 791-932, 791-973, 791-1004, 791-1007, 791-1009, 791-1014, 791-1023, 791-1048, 791-1057, 791-1063, 791-1080, 791-1096, 791-1101, 791-1131, 791-1132, 791-1186, 791-1215, 791-1225, 791-1235, 791-1248, 791-1266, 791-1348, 791-1402, 791-1416, 792-1237, 798-933, 798-1165, 800-1054, 800-1131,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
69	800-1357, 801-853, 801-985, 801-1100, 811-1015, 811-1318, 815-1249, 822-1216, 825-1351, 826-1040, 826-1063, 826-1107, 826-1477, 827-1319, 829-1309, 837-1225, 860-1225, 861-1443, 864-1274, 864-1459, 866-1309, 866-1351, 872-924, 872-929, 872-1008, 872-1082, 872-1091, 872-1102, 872-1133, 872-1142, 872-1185, 872-1237, 872-1274, 872-1291, 872-1432, 872-1439, 872-1500, 873-929, 874-1570, 875-924, 875-990, 875-1060, 875-1069, 875-1132, 875-1133, 875-1375, 875-1383, 879-1321, 882-998, 882-1012, 882-1191, 884-1112, 895-1402, 908-1299, 910-1333, 911-1403, 913-1357, 921-1309, 929-1309, 945-1531, 948-1357, 948-1551, 950-1357, 950-1435, 959-1105, 959-1175, 959-1226, 959-1231, 959-1266, 959-1584, 959-1654, 960-1405, 962-1561, 969-1222, 972-1008, 972-1237, 972-1340, 972-1345, 972-1357, 972-1551, 979-1486, 989-1384, 995-1417, 995-1487, 997-1477, 1005-1357, 1013-1357, 1026-1477, 1030-1477,
	1033-1520, 1034-1615, 1034-1622, 1038-1657, 1043-1093, 1043-1259, 1043-1266, 1043-1315, 1043-1351, 1043-1668, 1044-1489, 1046-1645, 1052-1092, 1052-1274, 1052-1522, 1052-1738, 1056-1092, 1056-1321, 1056-1424, 1056-1429, 1056-1627, 1063-1570, 1076-1468, 1078-1501, 1079-1571, 1089-1477, 1097-1477, 1107-1692, 1113-1699, 1117-1719, 1118-1561, 1121-1561, 1128-1573, 1137-1176, 1137-1357, 1137-1445, 1137-1606, 1137-1813, 1147-1561, 1148-1654, 1152-1343, 1152-1350, 1152-1399, 1152-1483, 1152-1752, 1157-1552, 1158-1250, 1158-1312, 1158-1585, 1158-1681, 1161-1839, 1163-1604, 1163-1655, 1173-1561, 1196-1561, 1197-1645, 1201-1645, 1202-1783, 1202-1803, 1204-1647, 1210-1334, 1210-1348, 1210-1396, 1210-1405, 1210-1508, 1210-1526, 1210-1528, 1210-1690, 1210-1719, 1211-1265, 1211-1418, 1211-1427, 1211-1432, 1211-1434, 1211-1480, 1211-1483, 1211-1489, 1211-1521, 1211-1592, 1211-1610, 1211-1777, 1211-1804, 1211-1811, 1213-1334, 1220-1474,
	1224-1611, 1231-1738, 1235-1669, 1242-1636, 1247-1739, 1249-1729, 1257-1645, 1277-1729, 1281-1813, 1284-1813, 1286-1729, 1286-1771, 1287-1645, 1294-1344, 1294-1432, 1294-1765, 1295-1813, 1304-1558, 1304-1741, 1304-1813, 1308-1344, 1308-1428, 1308-1502, 1308-1511, 1308-1516, 1308-1522, 1308-1564, 1308-1567, 1308-1573, 1308-1605, 1308-1676, 1308-1681, 1308-1695, 1308-1813, 1315-1813, 1325-1720, 1326-1753, 1329-1813, 1331-1777, 1333-1986, 1341-1729, 1349-1729, 1350-1986, 1354-1839, 1369-1787, 1370-1777, 1370-1813, 1379-1586, 1379-1595, 1379-1606, 1379-1645, 1379-1651, 1379-1735, 1379-1803, 1379-1813, 1380-1788, 1389-1642, 1392-1428, 1392-1657, 1392-1768, 1392-1796, 1392-1803, 1399-1777, 1409-1803, 1410-1813, 1413-1813, 1415-1813, 1417-1813, 1425-1966, 1433-2026, 1455-1813, 1456-1788, 1463-1502, 1463-1512, 1463-1600, 1463-1609, 1463-1670,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
69	1463-1679, 1463-1684, 1463-1690, 1463-1732, 1463-1735, 1463-1741, 1463-1765, 1463-1803, 1463-1813, 1464-1813, 1466-1502, 1472-1726, 1476-1813, 1483-1813, 1484-1803, 1492-1803, 1494-1813, 1499-1813, 1501-1813, 1509-1813, 1517-1813, 1533-1777, 1537-1813, 1538-1813, 1546-1596, 1546-1684, 1546-1813, 1547-1601, 1547-1754, 1547-1760, 1547-1763, 1547-1770, 1547-1813, 1548-1803, 1550-1813, 1556-1810, 1560-1596, 1567-1813, 1568-1777, 1577-1813, 1578-1813, 1581-1777, 1581-1813, 1583-1803, 1585-1813, 1593-1813, 1601-1813, 1617-1813, 1621-1813, 1622-1803, 1624-1795, 1631-1680, 1631-1690, 1631-1765, 1631-1777, 1631-1801, 1631-1802, 1631-1813, 1632-1777, 1640-1813, 1647-1680, 1651-1803, 1661-1813, 1662-1813, 1667-1813, 1669-1813, 1677-1813, 1698-1813, 1701-1813, 1705-1813, 1706-1813, 1708-1813, 1714-1764, 1714-1801, 1715-1801, 1716-1813, 1718-1813, 1728-1764, 1735-1813, 1739-1801, 1746-1777, 1751-1813, 1761-1813
70/7021892CB1/ 1724	1-1257, 780-1419, 780-1438, 780-1444, 780-1724

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
36	7492673CB1	BLADNOT03
37	7990930CB1	UTRSDIC01
38	7037554CB1	LUNGFEC01
39	1515347CB1	OVARNOT09
40	3464492CB1	UTRSNOT02
41	1794336CB1	THYMDIT01
42	2903694CB1	DRGCNOT01
43	6975426CB1	PROSTUS23
44	4019390CB1	BRABDIR03
45	986452CB1	THP1NOT03
46	2807579CB1	THP1AZT01
47	5724273CB1	MIXDUNB01
48	3614884CB1	EPIPNOT01
49	3794954CB1	PLACFER06
50	7399016CB1	SKINBIT01
51	6996690CB1	SINTNOR01
52	7740866CB1	LIVRTUE01
53	8181605CB1	BRAINOT03
54	8266487CB1	ADRENOT08
55	5552784CB1	SMCCNON03
56	7281230CB1	BMARTXE01
57	7488424CB1	BRAINOT19
58	7487110CB1	BRAWNOT01
59	7495008CB1	CORPNOT02
60	7073515CB1	BRAUTDR04
61	3356640CB1	BMARTXR02
62	2015706CB1	BRSTNOT02
63	6920755CB1	PLACFER06
64	444179CB1	MPHGNOT03
65	5628380CB1	PROSTUT09
66	7493789CB1	LIVRTUT01
67	2075194CB1	PGANNOT03
68	2801633CB1	BRADDIR01
69	7493525CB1	FTUBTUE01
70	7021892CB1	PANCNON03

Table 6

Library	Vector	Library Description
ADRENOT08	pINCY	Library was constructed using RNA isolated from adrenal tissue removed from a 20-year-old Caucasian male, who died from head trauma.
BLADNOT03	pINCY	Library was constructed using RNA isolated from bladder tissue removed from an 80-year-old Caucasian female during a radical cystectomy and lymph node excision. Pathology for the associated tumor tissue indicated grade 3 invasive transitional cell carcinoma. Patient history included malignant neoplasm of the uterus, atherosclerosis, and atrial fibrillation. Family history included acute renal failure, osteoarthritis, and atherosclerosis.
BMARTXE01	pINCY	This 5' biased random primed library was constructed using RNA isolated from treated SH-SY5Y cells derived from a metastatic bone marrow neuroblastoma, removed from a 4-year-old Caucasian female (Schering AG). The medium was MEM/HAM'S F12 with 10% fetal calf serum. After reaching about 80% confluency cells were treated with 6-Hydroxydopamine (6-OHDA) at 100 microM for 8 hours.
BMARTXR02	PCDNA2.1	This random primed library was constructed using RNA isolated from treated SH-SY5Y cells derived from a metastatic bone marrow neuroblastoma, removed from a 4-year-old Caucasian female (Schering AG). The medium was MEM/HAM'S F12 with 10% fetal calf serum. After reaching about 80% confluency cells were treated with 6-Hydroxydopamine (6-OHDA) at 100 microM for 8 hours.
BRABDIR03	pINCY	This random primed library was constructed using RNA isolated from diseased cerebellum tissue removed from the brain of a 57-year-old Caucasian male who died from a cerebrovascular accident. Serologies were negative. Patient history included Huntington's disease, emphysema, and tobacco abuse (3-4 packs per day for 40 years).
BRADDIR01	pINCY	Library was constructed using RNA isolated from diseased choroid plexus tissue of the lateral ventricle, removed from the brain of a 57-year-old Caucasian male, who died from a cerebrovascular accident.
BRAINOT03	PSPORT1	Library was constructed using RNA isolated from brain tissue removed from a 26-year-old Caucasian male during cranioplasty and excision of a cerebral meningeal lesion. Pathology for the associated tumor tissue indicated a grade 4 oligoastrocytoma in the right fronto-parietal part of the brain.

Table 6

Library	Vector	Library Description
BRAINOT19	pINCY	Library was constructed using RNA isolated from diseased brain tissue removed from the left frontal lobe of a 27-year-old Caucasian male during a brain lobectomy. Pathology indicated a focal deep white matter lesion, characterized by marked gliosis, calcifications, and hemosiderin-laden macrophages, consistent with a remote perinatal injury. This tissue also showed mild to moderate generalized gliosis, predominantly subpial and subcortical, consistent with chronic seizure disorder. The left temporal lobe, including the mesial temporal structures, showed focal, marked pyramidal cell loss and gliosis in hippocampal sector CA1, consistent with mesial temporal sclerosis. GFAP was positive for astrocytes. The patient presented with intractable epilepsy, focal epilepsy, hemiplegia, and an unspecified brain injury. Patient history included cerebral palsy, abnormality of gait, and depressive disorder. Family history included brain cancer.
BRAUTDR04	PCDNA2.1	This random primed library was constructed using RNA isolated from pooled striatum, dorsal caudate nucleus, dorsal putamen, and ventral nucleus accumbens tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.
BRAWNOT01	pINCY	Library was constructed using RNA isolated from dentate nucleus tissue removed from the brain of a 35-year-old Caucasian male who died from cardiac failure. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly, and an enlarged spleen and liver. Patient medications included simethicone, Lasix, Digoxin, Colace, Zantac, captopril, and Vasotec.
BRSTNOT02	PSPORT1	Library was constructed using RNA isolated from diseased breast tissue removed from a 55-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated proliferative fibrocysytic changes characterized by apocrine metaplasia, sclerosing adenosis, cyst formation, and ductal hyperplasia without atypia. Pathology for the associated tumor tissue indicated an invasive grade 4 mammary adenocarcinoma. Patient history included atrial tachycardia and a benign neoplasm. Family history included cardiovascular and cerebrovascular disease.
CORPNOT02	pINCY	Library was constructed using RNA isolated from diseased corpus callosum tissue removed from the brain of a 74-year-old Caucasian male who died from Alzheimer's disease.

Table 6

Library	Vector	Library Description
DRGCNOT01	pINCY	Library was constructed using RNA isolated from dorsal root ganglion tissue removed from the cervical spine of a 32-year-old Caucasian male who died from acute pulmonary edema and bronchopneumonia, bilateral pleural and pericardial effusions, and malignant lymphoma (natural killer cell type). Patient history included probable cytomegalovirus infection, hepatic congestion and steatosis, splenomegaly, hemorrhagic cystitis, thyroid hemorrhage, and Bell's palsy. Surgeries included colonoscopy, large intestine biopsy, adenotonsillectomy, and nasopharyngeal endoscopy and biopsy; treatment included radiation therapy.
EPIPNOT01	pINCY	Library was constructed using RNA isolated from prostatic epithelial cells removed from a 17-year-old Hispanic male.
FTUBTUE01	pINCY	This 5' biased random primed library was constructed using RNA isolated from right fallopian tube tumor tissue removed from an 85-year-old Caucasian female during bilateral salpingo-oophorectomy and hysterectomy. Pathology indicated poorly differentiated mixed endometrioid (80%) and serous (20%) adenocarcinoma of the right fallopian tube, which was confined to the mucosa without mural involvement. Endometrioid carcinoma in situ was also present. Pathology for the associated uterus tumor indicated focal endometrioid adenocarcinoma in situ and moderately differentiated invasive adenocarcinoma arising in an endometrial polyp. A metastatic endometrioid and serous adenocarcinoma was present in the cul-de-sac tumor. The patient presented with a pelvic mass and ascites. Patient history included medullary carcinoma of the thyroid and myocardial infarction. Patient medications included Nitro-Dur, Lescol, Lasix and Cardizem.
LIVRTUE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from liver tumor tissue removed from a 72-year-old Caucasian male during partial hepatectomy. Pathology indicated metastatic grade 2 (of 4) neuroendocrine carcinoma forming a mass. The patient presented with metastatic liver cancer. Patient history included benign hypertension, type I diabetes, prostatic hyperplasia, prostate cancer, alcohol abuse in remission, and tobacco abuse in remission. Previous surgeries included destruction of a pancreatic lesion, closed prostatic biopsy, transurethral prostatectomy, removal of bilateral testes and total splenectomy. Patient medications included Eulexin, Hytrin, Proscar, Ecotrin, and insulin. Family history included atherosclerotic coronary artery disease and acute myocardial infarction in the mother; atherosclerotic coronary artery disease and type II diabetes in the father.
LIVRTUT01	pINCY	Library was constructed using RNA isolated from liver tumor tissue removed from a 51-year-old Caucasian female during a hepatic lobectomy. Pathology indicated metastatic grade 3 adenocarcinoma consistent with colon cancer. Family history included a malignant neoplasm of the liver.
LUNGFEC01	pINCY	This large size-fractionated library was constructed using RNA isolated from lung tissue removed from a Caucasian male fetus who died from Patau's syndrome (trisomy 13) at 20-weeks' gestation.

Table 6

Library	Vector	Library Description
MIXDUNB01	pINCY	Library was constructed using RNA isolated from myometrium removed from a 41-year-old Caucasian female (A) during vaginal hysterectomy with a dilatation and curettage and untreated smooth muscle cells removed from the renal vein of a 57 year-old Caucasian male. Pathology for donor A indicated the myometrium and cervix were unremarkable. The endometrium was secretory and contained fragments of endometrial polyps. Benign endo- and ectocervical mucosa were identified in the endocervix. Pathology for the associated tumor tissue indicated uterine leiomyoma. Medical history included an unspecified menstrual disorder, ventral hernia, normal delivery, a benign ovarian neoplasm, and tobacco abuse in donor A. Previous surgeries included a bilateral destruction of fallopian tubes, removal of a solitary ovary, and an exploratory laparotomy in donor A. Medications included ferrous sulfate in donor A.
MPHGNOT03	PBLUESCRIPT	Library was constructed using RNA isolated from plastic adherent mononuclear cells isolated from buffy coat units obtained from unrelated male and female donors.
OVARNOT09	pINCY	Library was constructed using RNA isolated from ovarian tissue removed from a 28-year-old Caucasian female during a vaginal hysterectomy and removal of the fallopian tubes and ovaries. Pathology indicated multiple follicular cysts ranging in size from 0.4 to 1.5 cm in the right and left ovaries, chronic cervicitis and squamous metaplasia of the cervix, and endometrium in weakly proliferative phase. Family history included benign hypertension, hyperlipidemia, and atherosclerotic coronary artery disease.
PANCNON03	pINCY	This normalized pancreas tissue library was constructed from 12 million independent clones from a pancreas library. Starting RNA was made from RNA isolated from pancreas tissue removed from a 17-year-old Caucasian female who died from head trauma. Serology was positive for cytomegalovirus and remaining serologies were negative. The patient was not taking any medications. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
PGANNOT03	pINCY	Library was constructed using RNA isolated from paraganglionic tumor tissue removed from the intra-abdominal region of a 46-year-old Caucasian male during exploratory laparotomy. Pathology indicated a benign paraganglioma and was associated with a grade 2 renal cell carcinoma, clear cell type, which did not penetrate the capsule. Surgical margins were negative for tumor.
PLACFER06	pINCY	This random primed library was constructed using RNA isolated from placental tissue removed from a Caucasian fetus who died after 16 weeks' gestation from fetal demise and hydrocephalus. Patient history included umbilical cord wrapped around the head (3 times) and the shoulders (1 time). Serology was positive for anti-CMV. Family history included multiple pregnancies and live births, and an abortion.

Table 6

Library	Vector	Library Description
PROSTUS23	pINCY	This subtracted prostate tumor library was constructed using 10 million clones from a pooled prostate tumor library that was subjected to 2 rounds of subtractive hybridization with 10 million clones from a pooled prostate tissue library. The starting library for subtraction was constructed by pooling equal numbers of clones from 4 prostate tumor libraries using mRNA isolated from prostate tumor removed from Caucasian males at ages 58 (A), 61 (B), 66 (C), and 68 (D) during prostatectomy with lymph node excision. Pathology indicated adenocarcinoma in all donors. History included elevated PSA, induration and tobacco abuse in donor A; elevated PSA, induration, prostate hyperplasia, renal failure,
		osteoarthritis, renal artery stenosis, benign HTN, thrombocytopenia, hyperlipidemia, tobacco/alcohol abuse and hepatitis C (carrier) in donor B; elevated PSA, induration, and tobacco abuse in donor C; and elevated PSA, induration, hypercholesterolemia, and kidney calculus in donor D. The hybridization probe for subtraction was constructed by pooling equal numbers of cDNA clones from 3 prostate tissue libraries derived from prostate tissue, prostate epithelial cells, and fibroblasts from prostate stroma from 3 different donors. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR 19 (1991):1954 and Bonaldo, et al. Genome Research 6 (1996):791.
PROSTUT09	pINCY	Library was constructed using RNA isolated from prostate tumor tissue removed from a 66-year-old Caucasian male during a radical prostatectomy, radical cystectomy, and urinary diversion. Pathology indicated grade 3 transitional cell carcinoma. The patient presented with prostatic inflammatory disease. Patient history included lung neoplasm, and benign hypertension. Family history included a malignant breast neoplasm, tuberculosis, cerebrovascular disease, atherosclerotic coronary artery disease and lung cancer.
SINTNOR01	PCDNA2.1	This random primed library was constructed using RNA isolated from small intestine tissue removed from a 31-year-old Caucasian female during Roux-en-Y gastric bypass. Patient history included clinical obesity.
SKINBIT01	pINCY	Library was constructed using RNA isolated from diseased skin tissue of the left lower leg. Patient history included erythema nodosum of the left lower leg.
SMCCNON03	pINCY	This normalized smooth muscle cell library was constructed from 7.56 million independent clones from a smooth muscle cell library. Starting RNA was made from smooth muscle cell tissue removed from the coronary artery of a 3-year-old Caucasian male. The normalization and hybridization conditions were adapted from Soares et al., (PNAS (1994) 91:9228-9232); Swaroop et al., (NAR (1991) 19:1954); and Bonaldo et al., (Genome Research (1996) 6:791-806), using a significantly longer (48 hour) reannealing hybridization period.
THP1AZT01	pINCY	Library was constructed using RNA isolated from THP-1 promonocyte cells treated for three days with 0.8 micromolar 5-aza-2'-deoxycytidine. THP-1 (ATCC TIB 202) is a human promonocyte line derived from peripheral blood of a 1-year-old Caucasian male with acute monocytic leukemia (Int. J. Cancer (1980) 26:171).

Table 6

Library	Vector	Library Description
THP1NOT03	pINCY	Library was constructed using RNA isolated from untreated THP-1 cells. THP-1 is a human promonocyte line derived from the peripheral blood of a 1-year-old Caucasian male with acute monocytic leukemia (ref: Int. J. Cancer (1980) 26:171).
THYMDIT01	pINCY	The library was constructed using RNA isolated from diseased thymus tissue removed from a 16-year-old Caucasian female during a total excision of thymus and regional lymph node excision. Pathology indicated thymic follicular hyperplasia. The right lateral thymus showed reactive lymph nodes. A single reactive lymph node was also identified at the inferior thymus margin. The patient presented with myasthenia gravis, malaise, fatigue, dysphagia, severe muscle weakness and prominent eyes. Patient history included frozen face muscles. Family history included depressive disorder, hepatitis B, myocardial infarction, atherosclerotic coronary artery disease, leukemia, multiple sclerosis, and lupus.
UTRSDIC01	PSPORT1	This large size fractionated library was constructed using pooled cDNA from eight donors. cDNA was generated using mRNA isolated from endometrial tissue removed from a 32-year-old female (donor A); endometrial tissue removed from a 32-year-old Caucasian female (donor B) during abdominal hysterectomy, bilateral salpingo-oophorectomy, and cystocele repair; from diseased endometrium and myometrium tissue removed from a 38-year-old Caucasian female (donor C) during abdominal hysterectomy, bilateral salpingo-oophorectomy, and exploratory laparotomy; from endometrial tissue removed from a 41-year-old Caucasian female (donor D) during abdominal hysterectomy with removal of a solitary ovary;
		from endometrial tissue removed from a 43-year-old Caucasian female (donor E) during vaginal hysterectomy, dilation and curettage, cystocele repair, rectocele repair and cystostomy; and from endometrial tissue removed from a 48-year-old Caucasian female (donor F) during a vaginal hysterectomy, rectocele repair, and bilateral salpingo-oophorectomy. Pathology (A) indicated the endometrium was in secretory phase. Pathology (B) indicated the endometrium was in the proliferative phase. Pathology (C) indicated extensive adenomatous hyperplasia with squamous metaplasia and focal atypia, forming a polypoid mass within the endometrial cavity. The cervix showed chronic cervicitis and squamous metaplasia. Pathology (D, E) indicated the endometrium was secretory phase. Pathology (F) indicated the endometrium was weakly proliferative.
UTRSNOT02	PSPORT1	Library was constructed using RNA isolated from uterine tissue removed from a 34-year-old Caucasian female during a vaginal hysterectomy. Patient history included mitral valve disorder. Family history included stomach cancer, congenital heart anomaly, irritable bowel syndrome, ulcerative colitis, colon cancer, cerebrovascular disease, type II diabetes, and depression.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value=1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value=1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCY, SMART, and TIGRFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	*PFAM, INCY, SMART, or TIGRFAM hits:* Probability value=1.0E-3 or less *Signal peptide hits:* Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score=120 or greater; Match length=56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-35,

b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-2, SEQ ID NO:4-7, SEQ ID NO:9-16, SEQ ID NO:18-19, SEQ ID NO:21-22, SEQ ID NO:24, SEQ ID NO:27-35,

c) a polypeptide comprising a naturally occurring amino acid sequence at least 97% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:3, SEQ ID NO:17, and SEQ ID NO:25,

d) a polypeptide comprising a naturally occurring amino acid sequence at least 98% identical to the amino acid sequence of SEQ ID NO:8,

e) a polypeptide comprising a naturally occurring amino acid sequence at least 95% identical to the amino acid sequence of SEQ ID NO:20,

f) a polypeptide comprising a naturally occurring amino acid sequence at least 94% identical to the amino acid sequence of SEQ ID NO:23,

g) a polypeptide comprising a naturally occurring amino acid sequence at least 91% identical to the amino acid sequence of SEQ ID NO:26,

h) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, and

i) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-35.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-35.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:36-70.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-35.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:36-70,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:36-70,

c) a polynucleotide complementary to a polynucleotide of a),

d) a polynucleotide complementary to a polynucleotide of b), and

e) an RNA equivalent of a)-d).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides

comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-35.

19. A method for treating a disease or condition associated with decreased expression of functional NAAP, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of

functional NAAP, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional NAAP, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method

comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

a) treating a biological sample containing nucleic acids with the test compound,

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

c) quantifying the amount of hybridization complex, and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of NAAP in a biological sample, the method comprising:

a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of NAAP in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of NAAP in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibodies from the animal, and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-35.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-35, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-35.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-35 in a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-35 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-35 from a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-35.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

a) labeling the polynucleotides of the sample,

b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and

c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.

73. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.

74. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.

75. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.

76. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:21.

77. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:22.

78. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:23.

79. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:24.

80. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:25.

81. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:26.

82. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:27.

83. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:28.

84. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:29.

85. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:30.

86. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:31.

87. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:32.

88. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:33.

89. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:34.

90. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:35.

91. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID

NO:36.

92. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:37.

93. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:38.

94. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:39.

95. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:40.

96. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:41.

97. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:42.

98. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:43.

99. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:44.

100. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:45.

101. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:46.

102. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:47.

103. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:48.

104. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:49.

105. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:50.

106. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:51.

107. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:52.

108. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:53.

109. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:54.

110. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:55.

111. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:56.

112. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:57.

113. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:58.

114. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:59.

115. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:60.

116. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:61.

117. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:62.

118. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:63.

119. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:64.

120. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:65.

121. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:66.

122. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:67.

123. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:68.

124. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:69.

125. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:70.

<110> INCYTE GENOMICS, INC.
SWARNAKAR, Anita
RICHARDSON, Thomas W.
WARREN, Bridget A.
GRIFFIN, Jennifer A.
TANG, Y. Tom
YUE, Henry
BAUGHN, Mariah R.
EMERLING, Brooke M.
LAL, Preeti
LU, Dyung Aina M.
FORSYTHE, Ian J.
RAMKUMAR, Jayalaxmi
LI, Joana X.
BECHA, Shanya D.
DUGGAN, Brendan M.
SANJANWALA, Madhusudan M.
LEE, Ernestine A.
BURFORD, Neil
ELLIOTT, Vicki S.
ISON, Craig H.
DING, Li
BOROWSKY, Mark L.
YAO, Monique G.
BARROSO, Ines
TRAN, Bao
WALIA, Narinder K.
HAFALIA, April J. A.
NGUYEN, Danniel B.
LU, Yan
ARVIZU, Chandra

<120> NUCLEIC ACID-ASSOCIATED PROTEINS

<130> PF-1068 PCT

<140> To Be Assigned
<141> Herewith

<150> US 60/305,089
<151> 2001-07-12

<150> US 60/305,104
<151> 2001-07-12

<150> US 60/305,325
<151> 2001-07-13

<150> US 60/305,390
<151> 2001-07-13

<150> US 60/306,960
<151> 2001-07-19

<150> US 60/306,694
<151> 2001-07-20

<150> US 60/308,170
<151> 2001-07-27

<160> 70

<170> PERL Program

<210> 1
<211> 316
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7492673CD1

<400> 1
Met Ala Asp Asp Ala Ser Ala Glu Gly Gly Gly Gly Arg Ala Arg
1 5 10 15
Ala Gly Gly Arg Val Val Trp Ala Pro Glu Ala Pro Gly Ala Leu
20 25 30
Gly Arg Gly Thr Asp Val Thr Ser Ala Glu Gly Phe Ser Ser Gly
35 40 45
Ile Arg Val Arg Gly His Gly Arg Gly Trp Gly Arg Gly Arg Gly
50 55 60
Arg Gly Arg Gly Trp Gly Arg Gly Gln Asp Arg Gly Val Arg Gly
65 70 75
Gly Lys Ala Gln Asp Lys Glu Arg Met Pro Val Thr Lys Leu Asp
80 85 90
Gly Leu Ala Lys Asp Met Lys Ile Lys Ser Leu Glu Glu Ile Tyr
95 100 105
Leu Phe Ser Leu Pro Ile Lys Glu Ser Glu Ile Ile Asp Phe Phe
110 115 120
Leu Gly Ala Ser Leu Lys Asp Glu Val Leu Lys Ile Met Pro Val
125 130 135
Arg Lys Glu Thr Arg Ala Gly Gln Arg Thr Arg Phe Lys Ala Phe
140 145 150
Val Ala Ile Arg Asp Tyr Asn Gly Tyr Ala Gly Gly Leu Lys Cys
155 160 165
Ser Lys Glu Val Ala Ala Ala Ile Arg Gly Ala Ile Ile Leu Thr
170 175 180
Lys Leu Ser Ile Val Thr Val Arg Arg Gly Tyr Trp Gly Asn Lys
185 190 195
Ile Gly Lys Leu His Thr Val Pro Cys Lys Val Thr Gly Arg Cys
200 205 210
Gly Ser Val Leu Val Pro Phe Ile Pro Ala Pro Arg Gly Thr Gly
215 220 225
Ile Val Ser Ala Pro Val Pro Lys Lys Leu Leu Met Ile Ala Gly
230 235 240
Ile Tyr Asp Cys Tyr Thr Ser Ala Arg Gly Cys Thr Ala Thr Leu
245 250 255
Gly Asn Phe Ala Asn Ala Thr Leu Asp Ala Val Ser Asn Ile Tyr
260 265 270
Ser Tyr Leu Pro Leu Lys Asp Phe Trp Lys Glu Thr Val Phe Thr
275 280 285
Lys Ser Pro Tyr Gln Glu Phe Thr Asp His Leu Ala Lys Thr His
290 295 300
Thr Arg Val Ser Val Gln Arg Thr Gln Ala Pro Ala Val Thr Thr
305 310 315
Thr

<210> 2
<211> 192
<212> PRT
<213> Homo sapiens

<220>

<221> misc_feature
<223> Incyte ID No: 7990930CD1

<400> 2
Met Lys Ala Val Leu Ser Asn Gln Thr Val His Ile Pro Glu Asn
1 5 10 15
Val Asp Ile Thr Leu Lys Gly His Thr Val Leu Val Lys Gly Pro
20 25 30
Arg Gly Thr Leu Gln Arg Asp Phe Ser His Val Ser Ile Glu Leu
35 40 45
Gly Leu Leu Gly Lys Lys Lys Gln Arg Leu Gln Ile Asp Lys Cys
50 55 60
Met Leu Arg Arg Trp Glu Leu Ala Ala Met Arg Ala Ile Cys Ser
65 70 75
His Ala Gln Lys Met Ile Lys Gly Val Ile Leu Gly Phe Cys Asp
80 85 90
Lys Met Arg Ser Leu Cys Ser His Phe Pro Ile Asn Val Ile Met
95 100 105
Gln Glu Asn Ser Ser Ile Val Glu Val Arg Thr Phe Leu Ser Glu
110 115 120
Lys Tyr Ile Cys Arg Val Arg Val Ser Pro Gly Val Ala Cys Ser
125 130 135
Ile Ser Gln Ala Gln Lys Asp Glu Phe Ile Leu Lys Gly Asn Asp
140 145 150
Ile Glu Leu Val Ser Asn Ser Ala Ala Leu Thr Gln Gln Thr Thr
155 160 165
Thr Ala Lys Asn Gln Asp Ile Arg Thr Phe Leu Asp Gly Ile Tyr
170 175 180
Val Ser Glu Lys Gly Thr Val Gln Gln Ala Asp Glu
185 190

<210> 3
<211> 735
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7037554CD1

<400> 3
Met Ala Ala Asp Ser Arg Glu Glu Lys Asp Gly Glu Leu Asn Val
1 5 10 15
Leu Asp Asp Ile Leu Thr Glu Val Pro Glu Gln Asp Asp Glu Leu
20 25 30
Tyr Asn Pro Glu Ser Glu Gln Asp Lys Asn Glu Lys Lys Gly Ser
35 40 45
Lys Arg Lys Ser Asp Arg Met Glu Ser Thr Asp Thr Lys Arg Gln
50 55 60
Lys Pro Ser Val His Ser Arg Gln Leu Val Ser Lys Pro Leu Ser
65 70 75
Ser Ser Val Ser Asn Asn Lys Arg Ile Val Ser Thr Lys Gly Lys
80 85 90
Ser Ala Thr Glu Tyr Lys Asn Glu Glu Tyr Gln Arg Ser Glu Arg
95 100 105
Asn Lys Arg Leu Asp Ala Asp Arg Lys Ile Arg Leu Ser Ser Ser
110 115 120
Ala Ser Arg Glu Pro Tyr Lys Asn Gln Pro Glu Lys Thr Cys Val
125 130 135
Arg Lys Arg Asp Pro Glu Arg Arg Ala Lys Ser Pro Thr Pro Asp
140 145 150
Gly Ser Glu Arg Ile Gly Leu Glu Val Asp Arg Arg Ala Ser Arg
155 160 165

```
Ser Ser Gln Ser Ser Lys Glu Glu Val Asn Ser Glu Glu Tyr Gly
                170                 175                 180
Ser Asp His Glu Thr Gly Ser Ser Gly Ser Ser Asp Glu Gln Gly
                185                 190                 195
Asn Asn Thr Glu Asn Glu Glu Glu Gly Val Glu Glu Asp Val Glu
                200                 205                 210
Glu Asp Glu Glu Val Glu Glu Asp Ala Glu Glu Asp Glu Glu Val
                215                 220                 225
Asp Glu Asp Gly Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu Glu
                230                 235                 240
Glu Glu Glu Glu Glu Glu Glu Glu Glu Tyr Glu Gln Asp Glu Arg
                245                 250                 255
Asp Gln Lys Glu Glu Gly Asn Asp Tyr Asp Thr Arg Ser Glu Ala
                260                 265                 270
Ser Asp Ser Gly Ser Glu Ser Val Ser Phe Thr Asp Gly Ser Val
                275                 280                 285
Arg Ser Gly Ser Gly Thr Asp Gly Ser Asp Glu Lys Lys Lys Glu
                290                 295                 300
Arg Lys Arg Ala Arg Gly Ile Ser Pro Ile Val Phe Asp Arg Ser
                305                 310                 315
Gly Ser Ser Ala Ser Glu Ser Tyr Ala Gly Ser Glu Lys Lys His
                320                 325                 330
Glu Lys Leu Ser Ser Ser Val Arg Ala Val Arg Lys Asp Gln Thr
                335                 340                 345
Ser Lys Leu Lys Tyr Val Leu Gln Asp Ala Arg Phe Phe Leu Ile
                350                 355                 360
Lys Ser Asn Asn His Glu Asn Val Ser Leu Ala Lys Ala Lys Gly
                365                 370                 375
Val Trp Ser Thr Leu Pro Val Asn Glu Lys Lys Leu Asn Leu Ala
                380                 385                 390
Phe Arg Ser Ala Arg Ser Val Ile Leu Ile Phe Ser Val Arg Glu
                395                 400                 405
Ser Gly Lys Phe Gln Gly Phe Ala Arg Leu Ser Ser Glu Ser His
                410                 415                 420
His Gly Gly Ser Pro Ile His Trp Val Leu Pro Ala Gly Met Ser
                425                 430                 435
Ala Lys Met Leu Gly Gly Val Phe Lys Ile Asp Trp Ile Cys Arg
                440                 445                 450
Arg Glu Leu Pro Phe Thr Lys Ser Ala His Leu Thr Asn Pro Trp
                455                 460                 465
Asn Glu His Lys Pro Val Lys Ile Gly Arg Asp Gly Gln Glu Ile
                470                 475                 480
Glu Leu Glu Cys Gly Thr Gln Leu Cys Leu Leu Phe Pro Pro Asp
                485                 490                 495
Glu Ser Ile Asp Leu Tyr Gln Val Ile His Lys Met Arg His Lys
                500                 505                 510
Arg Arg Met His Ser Gln Pro Arg Ser Arg Gly Arg Pro Ser Arg
                515                 520                 525
Arg Glu Pro Val Arg Asp Val Gly Arg Arg Arg Pro Glu Asp Tyr
                530                 535                 540
Asp Ile His Asn Ser Arg Lys Lys Pro Arg Ile Asp Tyr Pro Pro
                545                 550                 555
Glu Phe His Gln Arg Pro Gly Tyr Leu Lys Asp Pro Arg Tyr Gln
                560                 565                 570
Glu Val Asp Ser Phe Thr Asn Leu Ile Pro Asn Arg Arg Phe Ser
                575                 580                 585
Gly Val Arg Arg Asp Val Phe Leu Asn Gly Ser Tyr Asn Asp Tyr
                590                 595                 600
Val Arg Glu Phe His Asn Met Gly Pro Pro Pro Pro Trp Gln Gly
                605                 610                 615
Met Pro Pro Tyr Pro Gly Met Glu Gln Pro Pro His His Pro Tyr
                620                 625                 630
Tyr Gln His His Ala Pro Pro Pro Gln Ala His Pro Pro Tyr Ser
```

```
                635                 640                 645
Gly His His Pro Val Pro His Glu Ala Arg Tyr Arg Asp Lys Arg
                650                 655                 660
Val His Asp Tyr Asp Met Arg Val Asp Asp Phe Leu Arg Arg Thr
                665                 670                 675
Gln Ala Val Val Ser Gly Arg Arg Ser Arg Pro Arg Glu Arg Asp
                680                 685                 690
Arg Glu Arg Glu Arg Asp Arg Pro Arg Asp Asn Arg Arg Asp Arg
                695                 700                 705
Glu Arg Asp Arg Gly Arg Asp Arg Glu Arg Glu Arg Glu Arg Leu
                710                 715                 720
Cys Asp Arg Asp Arg Asp Arg Gly Glu Arg Gly Arg Tyr Arg Arg
                725                 730                 735


<210> 4
<211> 1340
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1515347CD1

<400> 4
Met Leu Thr Leu Cys Arg Cys Gly Glu Ser Leu Gln Asp Val Ile
  1                 5                  10                  15
Asp Arg Val Ala Phe Val Ile Pro Pro Val Val Ala Ala Pro Pro
                 20                  25                  30
Ser Leu Arg Val Pro Arg Pro Pro Pro Leu Tyr Ser His Arg Met
                 35                  40                  45
Arg Ile Leu Arg Gln Gly Leu Arg Glu His Ala Ala Pro Tyr Phe
                 50                  55                  60
Gln Gln Leu Arg Gln Thr Thr Ala Pro Arg Leu Leu Gln Phe Pro
                 65                  70                  75
Glu Leu Arg Leu Val Gln Phe Asp Ser Gly Lys Leu Glu Ala Leu
                 80                  85                  90
Ala Ile Leu Leu Gln Lys Leu Lys Ser Glu Gly Arg Arg Val Leu
                 95                 100                 105
Ile Leu Ser Gln Met Ile Leu Met Leu Asp Ile Leu Glu Met Phe
                110                 115                 120
Leu Asn Phe His Tyr Leu Thr Tyr Val Arg Ile Asp Glu Asn Ala
                125                 130                 135
Ser Ser Glu Gln Arg Gln Glu Leu Met Arg Ser Phe Asn Arg Asp
                140                 145                 150
Arg Arg Ile Phe Cys Ala Ile Leu Ser Thr His Ser Arg Thr Thr
                155                 160                 165
Gly Ile Asn Leu Val Glu Ala Asp Thr Val Val Phe Tyr Asp Asn
                170                 175                 180
Asp Leu Asn Pro Val Met Asp Ala Lys Ala Gln Glu Trp Cys Asp
                185                 190                 195
Arg Ile Gly Arg Cys Lys Asp Ile His Ile Tyr Arg Leu Val Ser
                200                 205                 210
Gly Asn Ser Ile Glu Glu Lys Leu Leu Lys Asn Gly Thr Lys Asp
                215                 220                 225
Leu Ile Arg Glu Val Ala Ala Gln Gly Asn Asp Tyr Ser Met Ala
                230                 235                 240
Phe Leu Thr Gln Arg Thr Ile Gln Glu Leu Phe Glu Val Tyr Ser
                245                 250                 255
Pro Met Asp Asp Ala Gly Phe Pro Val Lys Ala Glu Glu Phe Val
                260                 265                 270
Val Leu Ser Gln Glu Pro Ser Val Thr Glu Thr Ile Ala Pro Lys
                275                 280                 285
```

```
Ile Ala Arg Pro Phe Ile Glu Ala Leu Lys Ser Ile Glu Tyr Leu
                290                 295                 300
Glu Glu Asp Ala Gln Lys Ser Ala Gln Glu Gly Val Leu Gly Pro
                305                 310                 315
His Thr Asp Ala Leu Ser Ser Asp Ser Glu Asn Met Pro Cys Asp
                320                 325                 330
Glu Glu Pro Ser Gln Leu Glu Glu Leu Ala Asp Phe Met Glu Gln
                335                 340                 345
Leu Thr Pro Ile Glu Lys Tyr Ala Leu Asn Tyr Leu Glu Leu Phe
                350                 355                 360
His Thr Ser Ile Glu Gln Glu Lys Glu Arg Asn Ser Glu Asp Ala
                365                 370                 375
Val Met Thr Ala Val Arg Ala Trp Glu Phe Trp Asn Leu Lys Thr
                380                 385                 390
Leu Gln Glu Arg Glu Ala Arg Leu Arg Leu Glu Gln Glu Glu Ala
                395                 400                 405
Glu Leu Leu Thr Tyr Thr Arg Glu Asp Ala Tyr Ser Met Glu Tyr
                410                 415                 420
Val Tyr Glu Asp Val Asp Gly Gln Thr Glu Val Met Pro Leu Trp
                425                 430                 435
Thr Pro Pro Thr Pro Pro Gln Asp Asp Ser Asp Ile Tyr Leu Asp
                440                 445                 450
Ser Val Met Cys Leu Met Tyr Glu Ala Thr Pro Ile Pro Glu Ala
                455                 460                 465
Lys Leu Pro Pro Val Tyr Val Arg Lys Glu Arg Lys Arg His Lys
                470                 475                 480
Thr Asp Pro Ser Ala Ala Gly Arg Lys Lys Lys Gln Arg His Gly
                485                 490                 495
Glu Ala Val Val Pro Pro Arg Ser Leu Phe Asp Arg Ala Thr Pro
                500                 505                 510
Gly Leu Leu Lys Ile Arg Arg Glu Gly Lys Glu Gln Lys Lys Asn
                515                 520                 525
Ile Leu Leu Lys Gln Gln Val Pro Phe Ala Lys Pro Leu Pro Thr
                530                 535                 540
Phe Ala Lys Pro Thr Ala Glu Pro Gly Gln Asp Asn Pro Glu Trp
                545                 550                 555
Leu Ile Ser Glu Asp Trp Ala Leu Leu Gln Ala Val Lys Gln Leu
                560                 565                 570
Leu Glu Leu Pro Leu Asn Leu Thr Ile Val Ser Pro Ala His Thr
                575                 580                 585
Pro Asn Trp Asp Leu Val Ser Asp Val Val Asn Ser Cys Ser Arg
                590                 595                 600
Ile Tyr Arg Ser Ser Lys Gln Cys Arg Asn Arg Tyr Glu Asn Val
                605                 610                 615
Ile Ile Pro Arg Glu Glu Gly Lys Ser Lys Asn Asn Arg Pro Leu
                620                 625                 630
Arg Thr Ser Gln Ile Tyr Ala Gln Asp Glu Asn Ala Thr His Thr
                635                 640                 645
Gln Leu Tyr Thr Ser His Phe Asp Leu Met Lys Met Thr Ala Gly
                650                 655                 660
Lys Arg Ser Pro Pro Ile Lys Pro Leu Leu Gly Met Asn Pro Phe
                665                 670                 675
Gln Lys Asn Pro Lys His Ala Ser Val Leu Ala Glu Ser Gly Ile
                680                 685                 690
Asn Tyr Asp Lys Pro Leu Pro Pro Ile Gln Val Ala Ser Leu Arg
                695                 700                 705
Ala Glu Arg Ile Ala Lys Glu Lys Lys Ala Leu Ala Asp Gln Gln
                710                 715                 720
Lys Ala Gln Gln Pro Ala Val Ala Gln Pro Pro Pro Pro Gln Pro
                725                 730                 735
Gln Pro Pro Pro Pro Pro Gln Gln Pro Pro Pro Pro Leu Pro Gln
                740                 745                 750
Pro Gln Ala Ala Gly Ser Gln Pro Pro Ala Gly Pro Pro Ala Val
```

```
                 755                 760                 765
Gln Pro Gln Pro Gln Pro Gln Pro Gln Thr Gln Pro Gln Pro Val
                 770                 775                 780
Gln Ala Pro Ala Lys Ala Gln Pro Ala Ile Thr Thr Gly Gly Ser
                 785                 790                 795
Ala Ala Val Leu Ala Gly Thr Ile Lys Thr Ser Val Thr Gly Thr
                 800                 805                 810
Ser Met Pro Thr Gly Ala Val Ser Gly Asn Val Ile Val Asn Thr
                 815                 820                 825
Ile Ala Gly Val Pro Ala Ala Thr Phe Gln Ser Ile Asn Lys Arg
                 830                 835                 840
Leu Ala Ser Pro Val Ala Pro Gly Ala Leu Thr Thr Pro Gly Gly
                 845                 850                 855
Ser Ala Pro Ala Gln Val Val His Thr Gln Pro Pro Pro Arg Ala
                 860                 865                 870
Val Gly Ser Pro Ala Thr Ala Thr Pro Asp Leu Val Ser Met Ala
                 875                 880                 885
Thr Thr Gln Gly Val Arg Ala Val Thr Ser Val Thr Ala Ser Ala
                 890                 895                 900
Val Val Thr Thr Asn Leu Thr Pro Val Gln Thr Pro Ala Arg Ser
                 905                 910                 915
Leu Val Pro Gln Val Ser Gln Ala Thr Gly Val Gln Leu Pro Gly
                 920                 925                 930
Lys Thr Ile Thr Pro Ala His Phe Gln Leu Leu Arg Gln Gln Gln
                 935                 940                 945
Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln
                 950                 955                 960
Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Thr Thr Thr
                 965                 970                 975
Thr Ser Gln Val Gln Val Pro Gln Ile Gln Gly Gln Ala Gln Ser
                 980                 985                 990
Pro Ala Gln Ile Lys Ala Val Gly Lys Leu Thr Pro Glu His Leu
                 995                1000                1005
Ile Lys Met Gln Lys Gln Lys Leu Gln Met Pro Pro Gln Pro Pro
                1010                1015                1020
Pro Pro Gln Ala Gln Ser Ala Pro Pro Gln Pro Thr Ala Gln Val
                1025                1030                1035
Gln Val Gln Thr Ser Gln Pro Pro Gln Gln Gln Ser Pro Gln Leu
                1040                1045                1050
Thr Thr Val Thr Ala Pro Arg Pro Gly Ala Leu Leu Thr Gly Thr
                1055                1060                1065
Thr Val Ala Asn Leu Gln Val Ala Arg Leu Leu Gln Ala Gln Gly
                1070                1075                1080
Gln Met Gln Thr Gln Ala Pro Gln Pro Ala Gln Val Ala Leu Ala
                1085                1090                1095
Lys Pro Pro Val Val Ser Val Pro Ala Ala Val Val Ser Ser Pro
                1100                1105                1110
Gly Val Thr Thr Leu Pro Met Asn Val Ala Gly Ile Ser Val Ala
                1115                1120                1125
Ile Gly Gln Pro Gln Lys Ala Ala Gly Gln Thr Val Val Ala Gln
                1130                1135                1140
Ala Arg His Met Gln Gln Leu Leu Lys Leu Lys Gln Gln Ala Val
                1145                1150                1155
Gln Gln Gln Lys Ala Ile Gln Pro Gln Ala Ala Gln Gly Pro Ala
                1160                1165                1170
Ala Val Gln Gln Lys Ile Thr Ala Gln Gln Ile Thr Thr Pro Gly
                1175                1180                1185
Ala Gln Gln Lys Val Ala Tyr Ala Ala Gln Pro Ala Leu Lys Thr
                1190                1195                1200
Gln Phe Leu Thr Thr Pro Ile Ser Gln Ala Gln Lys Leu Ala Gly
                1205                1210                1215
Ala Gln Gln Val Gln Thr Gln Ile Gln Val Ala Lys Leu Pro Gln
                1220                1225                1230
```

```
Val Val Gln Gln Gln Thr Pro Val Ala Ser Ile Gln Gln Val Ala
                1235                1240                1245
Ser Ala Ser Gln Gln Ala Ser Pro Gln Thr Val Ala Leu Thr Gln
                1250                1255                1260
Ala Thr Ala Ala Gly Gln Gln Val Gln Met Ile Pro Ala Val Thr
                1265                1270                1275
Ala Thr Ala Gln Val Val Gln Gln Lys Leu Ile Gln Gln Gln Val
                1280                1285                1290
Val Thr Thr Ala Ser Ala Pro Leu Gln Thr Pro Gly Ala Pro Asn
                1295                1300                1305
Pro Ala Gln Val Pro Ala Ser Ser Asp Ser Pro Ser Gln Gln Pro
                1310                1315                1320
Lys Leu Gln Met Arg Val Pro Ala Val Arg Leu Lys Thr Pro Thr
                1325                1330                1335
Lys Pro Pro Cys Gln
                1340
```

<210> 5
<211> 560
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3464492CD1

<400> 5

```
Met Thr Phe Ser Arg Leu Leu Asn Tyr Lys Tyr Ser Asp Thr Leu
  1                 5                  10                  15
Lys Lys Met Asp Pro Asp His Leu Val Ala Leu Val Thr Glu Val
                 20                  25                  30
Ile Pro Asn Tyr Ser Cys Leu Val Phe Cys Pro Ser Lys Lys Asn
                 35                  40                  45
Cys Glu Asn Val Ala Glu Met Ile Cys Lys Phe Leu Ser Lys Glu
                 50                  55                  60
Tyr Leu Lys His Lys Glu Lys Glu Lys Cys Glu Val Ile Lys Asn
                 65                  70                  75
Leu Lys Asn Ile Gly Asn Gly Asn Leu Cys Pro Val Leu Lys Arg
                 80                  85                  90
Thr Ile Pro Phe Gly Val Ala Tyr His His Ser Gly Leu Thr Ser
                 95                 100                 105
Asp Glu Arg Lys Leu Leu Glu Glu Ala Tyr Ser Thr Gly Val Leu
                110                 115                 120
Cys Leu Phe Thr Cys Thr Ser Thr Leu Ala Ala Gly Val Asn Leu
                125                 130                 135
Pro Ala Arg Arg Val Ile Leu Arg Ala Pro Tyr Val Ala Lys Glu
                140                 145                 150
Phe Leu Lys Arg Asn Gln Tyr Lys Gln Met Ile Gly Arg Ala Gly
                155                 160                 165
Arg Ala Gly Ile Asp Thr Ile Gly Glu Ser Ile Leu Ile Leu Gln
                170                 175                 180
Glu Lys Asp Lys Gln Gln Val Leu Glu Leu Ile Thr Lys Pro Leu
                185                 190                 195
Glu Asn Cys Tyr Ser His Leu Val Gln Glu Phe Thr Lys Gly Ile
                200                 205                 210
Gln Thr Leu Phe Leu Ser Leu Ile Gly Leu Lys Ile Ala Thr Asn
                215                 220                 225
Leu Asp Asp Ile Tyr His Phe Met Asn Gly Thr Phe Phe Gly Val
                230                 235                 240
Gln Gln Lys Val Leu Leu Lys Glu Lys Ser Leu Trp Glu Ile Thr
                245                 250                 255
Val Glu Ser Leu Arg Tyr Leu Thr Glu Lys Gly Leu Leu Gln Lys
                260                 265                 270
```

```
Asp Thr Ile Tyr Lys Ser Glu Glu Glu Val Gln Tyr Asn Phe His
                  275                 280                 285
Ile Thr Lys Leu Gly Arg Ala Ser Phe Lys Gly Thr Ile Asp Leu
                  290                 295                 300
Ala Tyr Cys Asp Ile Leu Tyr Arg Asp Leu Lys Lys Gly Leu Glu
                  305                 310                 315
Gly Leu Val Leu Glu Ser Leu Leu His Leu Ile Tyr Leu Thr Thr
                  320                 325                 330
Pro Tyr Asp Leu Val Ser Gln Cys Asn Pro Asp Trp Met Ile Tyr
                  335                 340                 345
Phe Arg Gln Phe Ser Gln Leu Ser Pro Ala Glu Gln Asn Val Ala
                  350                 355                 360
Ala Ile Leu Gly Val Ser Glu Ser Phe Ile Gly Lys Lys Ala Ser
                  365                 370                 375
Gly Gln Ala Ile Gly Lys Lys Val Asp Lys Asn Val Val Asn Arg
                  380                 385                 390
Leu Tyr Leu Ser Phe Val Leu Tyr Thr Leu Leu Lys Glu Thr Asn
                  395                 400                 405
Ile Trp Thr Val Ser Glu Lys Phe Asn Met Pro Arg Gly Tyr Ile
                  410                 415                 420
Gln Asn Leu Leu Thr Gly Thr Ala Ser Phe Ser Ser Cys Val Leu
                  425                 430                 435
His Phe Cys Glu Glu Leu Glu Glu Phe Trp Val Tyr Arg Ala Leu
                  440                 445                 450
Leu Val Glu Leu Thr Lys Lys Leu Thr Tyr Cys Val Lys Ala Glu
                  455                 460                 465
Leu Ile Pro Leu Met Glu Val Thr Gly Val Leu Glu Gly Arg Ala
                  470                 475                 480
Lys Gln Leu Tyr Ser Ala Gly Tyr Lys Ser Leu Met His Leu Ala
                  485                 490                 495
Asn Ala Asn Pro Glu Val Leu Val Arg Thr Ile Asp His Leu Ser
                  500                 505                 510
Arg Arg Gln Ala Lys Gln Ile Val Ser Ser Ala Lys Met Leu Leu
                  515                 520                 525
His Glu Lys Ala Glu Ala Leu Gln Glu Glu Val Glu Glu Leu Leu
                  530                 535                 540
Arg Leu Pro Ser Asp Phe Leu Val Leu Trp Leu Leu Pro Leu Thr
                  545                 550                 555
Lys His Glu Ala Ile
                  560
```

<210> 6
<211> 436
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1794336CD1

<400> 6

```
Met Glu Glu Phe Lys Ser His Ser Pro Glu Arg Ser Ile Phe Ser
  1                 5                  10                  15
Ala Ile Trp Glu Gly Asn Cys His Phe Glu Gln His Gln Gly Gln
                  20                  25                  30
Glu Glu Gly Tyr Phe Arg Gln Leu Met Ile Asn His Glu Asn Met
                  35                  40                  45
Pro Ile Phe Ser Gln His Thr Leu Leu Thr Gln Glu Phe Tyr Asp
                  50                  55                  60
Arg Glu Lys Ile Ser Glu Cys Lys Lys Cys Arg Lys Ile Phe Ser
                  65                  70                  75
Tyr His Leu Phe Phe Ser His His Lys Arg Thr His Ser Lys Glu
                  80                  85                  90
```

```
Leu Ser Glu Cys Lys Glu Cys Thr Glu Ile Val Asn Thr Pro Cys
                 95                  100                 105
Leu Phe Lys Gln Gln Thr Ile Gln Asn Gly Asp Lys Cys Asn Glu
                110                 115                 120
Cys Lys Glu Cys Trp Lys Ala Phe Val His Cys Ser His Phe Lys
                125                 130                 135
His Leu Arg Ile His Asn Gly Glu Lys Arg Tyr Glu Cys Asn Glu
                140                 145                 150
Cys Gly Lys Ala Phe Asn Tyr Gly Ser Glu Leu Thr Leu His Gln
                155                 160                 165
Arg Ile His Thr Gly Glu Lys Pro Tyr Glu Cys Lys Glu Cys Gly
                170                 175                 180
Lys Ala Phe Arg Gln Arg Ser Gln Leu Thr Gln His Gln Arg Leu
                185                 190                 195
His Thr Gly Glu Lys Pro Tyr Glu Cys Lys Gln Cys Gly Lys Ala
                200                 205                 210
Phe Ile Arg Gly Phe Gln Leu Thr Glu His Leu Arg Leu His Thr
                215                 220                 225
Gly Glu Lys Pro Tyr Glu Cys Lys Glu Cys Gly Lys Thr Phe Arg
                230                 235                 240
His Arg Ser His Leu Thr Ile His Gln Arg Ile His Thr Gly Glu
                245                 250                 255
Lys Pro Tyr Glu Cys Arg Glu Cys Gly Lys Ala Phe Ser Tyr His
                260                 265                 270
Ser Ser Phe Ser His His Gln Lys Ile His Ser Gly Lys Lys Pro
                275                 280                 285
Tyr Glu Cys His Glu Cys Gly Lys Ala Phe Cys Asp Gly Leu Gln
                290                 295                 300
Leu Thr Leu His Gln Arg Ile His Thr Gly Glu Lys Pro Tyr Glu
                305                 310                 315
Cys Lys Glu Cys Gly Lys Thr Phe Arg Gln Cys Ser His Leu Lys
                320                 325                 330
Arg His Gln Arg Ile His Thr Gly Glu Lys Pro His Glu Cys Met
                335                 340                 345
Ile Cys Gly Lys Ala Phe Arg Leu His Ser His Leu Ile Gln His
                350                 355                 360
Gln Arg Ile His Thr Gly Glu Lys Pro Tyr Glu Cys Lys Glu Cys
                365                 370                 375
Gly Lys Ala Phe Ser Tyr His Ser Ser Phe Ser His His Gln Arg
                380                 385                 390
Ile His Ser Gly Lys Lys Pro Tyr Gln Cys Gly Lys Ala Phe Asn
                395                 400                 405
His Arg Leu Gln Leu Asn Leu His Gln Thr Leu His Thr Gly Glu
                410                 415                 420
Lys Pro Val Arg Phe Pro Leu Leu Pro Pro His Pro Ser Leu Ala
                425                 430                 435
Ser
```

<210> 7
<211> 817
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2903694CD1

<400> 7

```
Met Asp Arg Asp Leu Glu Gln Ala Leu Asp Arg Thr Glu Asn Ile
  1               5                  10                  15
Thr Glu Ile Ala Gln Gln Arg Arg Pro Arg Arg Arg Tyr Ser Pro
                 20                  25                  30
```

```
Arg Ala Gly Lys Thr Leu Gln Glu Lys Leu Tyr Asp Ile Tyr Val
                35                  40                  45
Glu Glu Cys Gly Lys Glu Pro Glu Asp Pro Gln Glu Leu Arg Ser
            50                  55                  60
Asn Val Asn Leu Leu Glu Lys Leu Val Arg Arg Glu Ser Leu Pro
        65                  70                  75
Cys Leu Leu Val Asn Leu Tyr Pro Gly Asn Gln Gly Tyr Ser Val
    80                  85                  90
Met Leu Gln Arg Glu Asp Gly Ser Phe Ala Glu Thr Ile Arg Leu
            95                  100                 105
Pro Tyr Glu Glu Arg Ala Leu Leu Asp Tyr Leu Asp Ala Glu Glu
        110                 115                 120
Leu Pro Pro Ala Leu Gly Asp Val Leu Asp Lys Ala Ser Val Asn
    125                 130                 135
Ile Phe His Ser Gly Cys Val Ile Val Glu Val Arg Asp Tyr Arg
140                 145                 150
Gln Ser Ser Asn Met Gln Pro Pro Gly Tyr Gln Ser Arg His Ile
155                 160                 165
Leu Leu Arg Pro Thr Met Gln Thr Leu Ala Pro Glu Val Lys Thr
170                 175                 180
Met Thr Arg Asp Gly Glu Lys Trp Ser Gln Glu Asp Lys Phe Pro
185                 190                 195
Leu Glu Ser Gln Leu Ile Leu Ala Thr Ala Glu Pro Leu Cys Leu
200                 205                 210
Asp Pro Ser Val Ala Val Ala Cys Thr Ala Asn Arg Leu Leu Tyr
215                 220                 225
Asn Lys Gln Lys Met Asn Thr Asp Pro Met Glu Gln Cys Leu Gln
230                 235                 240
Arg Tyr Ser Trp Pro Ser Val Lys Pro Gln Gln Glu Gln Ser Asp
245                 250                 255
Cys Pro Pro Pro Pro Glu Leu Arg Val Ser Thr Ser Gly Gln Lys
260                 265                 270
Glu Glu Arg Lys Val Gly Gln Pro Cys Glu Leu Asn Ile Thr Lys
275                 280                 285
Ala Gly Ser Cys Val Asp Thr Trp Lys Gly Arg Pro Cys Asp Leu
290                 295                 300
Ala Val Pro Ser Glu Val Asp Val Glu Lys Leu Ala Lys Gly Tyr
305                 310                 315
Gln Ser Val Thr Ala Ala Asp Pro Gln Leu Pro Val Trp Pro Ala
320                 325                 330
Gln Glu Val Glu Asp Pro Phe Arg His Ala Trp Glu Ala Gly Cys
335                 340                 345
Gln Ala Trp Asp Thr Lys Pro Asn Ile Met Gln Ser Phe Asn Asp
350                 355                 360
Pro Leu Leu Cys Gly Lys Ile Arg Pro Arg Lys Lys Ala Arg Gln
365                 370                 375
Lys Ser Gln Lys Ser Pro Trp Gln Pro Phe Pro Asp Asp His Ser
380                 385                 390
Ala Cys Leu Arg Pro Gly Ser Glu Thr Asp Ala Gly Arg Ala Val
395                 400                 405
Ser Gln Ala Gln Glu Ser Val Gln Ser Lys Val Lys Gly Pro Gly
410                 415                 420
Lys Met Ser His Ser Ser Ser Gly Pro Ala Ser Val Ser Gln Leu
425                 430                 435
Ser Ser Trp Lys Thr Pro Glu Gln Pro Asp Pro Val Trp Val Gln
440                 445                 450
Ser Ser Val Ser Gly Lys Gly Glu Lys His Pro Pro Pro Arg Thr
455                 460                 465
Gln Leu Pro Ser Ser Ser Gly Lys Ile Ser Ser Gly Asn Ser Phe
470                 475                 480
Pro Pro Gln Gln Ala Gly Ser Pro Leu Lys Arg Pro Phe Pro Ala
485                 490                 495
Ala Ala Pro Ala Val Ala Ala Ala Ala Pro Ala Pro Ala Pro Ala
```

```
                500                 505                 510
Pro Ala Ala Ala Pro Ala Leu Ala Ala Ala Ala Val Ala Ala Ala
                515                 520                 525
Ala Gly Gly Ala Ala Pro Ser His Ser Gln Lys Pro Ser Val Pro
                530                 535                 540
Leu Ile Lys Ala Ser Arg Arg Arg Pro Ala Ala Gly Arg Pro Thr
                545                 550                 555
Arg Phe Val Lys Ile Ala Pro Ala Ile Gln Val Arg Thr Gly Ser
                560                 565                 570
Thr Gly Leu Lys Ala Thr Asn Val Glu Gly Pro Val Arg Gly Ala
                575                 580                 585
Gln Val Leu Gly Cys Ser Phe Lys Pro Val Gln Ala Pro Gly Ser
                590                 595                 600
Gly Ala Pro Ala Pro Ala Gly Ile Ser Gly Ser Gly Leu Gln Ser
                605                 610                 615
Ser Gly Gly Pro Leu Pro Asp Ala Arg Pro Gly Ala Val Gln Ala
                620                 625                 630
Ser Ser Pro Ala Pro Leu Gln Phe Phe Leu Asn Thr Pro Glu Gly
                635                 640                 645
Leu Arg Pro Leu Thr Leu Gln Val Pro Gln Gly Trp Ala Val Leu
                650                 655                 660
Thr Gly Pro Gln Gln Gln Ser His Gln Leu Val Ser Leu Gln Gln
                665                 670                 675
Leu Gln Gln Pro Thr Ala Ala His Pro Pro Gln Pro Gly Pro Gln
                680                 685                 690
Gly Ser Thr Leu Gly Leu Ser Thr Gln Gly Gln Ala Phe Pro Ala
                695                 700                 705
Gln Gln Leu Leu Asn Val Asn Leu Thr Gly Ala Gly Ser Gly Leu
                710                 715                 720
Gln Pro Gln Pro Gln Ala Ala Val Leu Ser Leu Leu Gly Ser Ala
                725                 730                 735
Gln Val Pro Gln Gln Gly Val Gln Leu Pro Phe Val Leu Gly Gln
                740                 745                 750
Gln Pro Gln Pro Leu Leu Leu Leu Gln Pro Gln Pro Gln Pro Gln
                755                 760                 765
Gln Ile Gln Leu Gln Thr Gln Pro Leu Arg Val Leu Gln Gln Pro
                770                 775                 780
Val Phe Leu Ala Thr Gly Ala Val Gln Ile Val Gln Pro His Pro
                785                 790                 795
Gly Val Gln Ala Gly Ser Gln Leu Val Gly Gln Arg Lys Gly Gly
                800                 805                 810
Lys Pro Thr Pro Pro Ala Pro
                815
```

<210> 8
<211> 1273
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6975426CD1

<400> 8

```
Met Lys Arg Arg Leu Asp Asp Gln Glu Ser Pro Val Tyr Ala Ala
  1               5                  10                  15
Gln Gln Arg Arg Ile Pro Gly Ser Thr Glu Ala Phe Pro His Gln
                 20                  25                  30
His Arg Val Leu Ala Pro Ala Pro Pro Val Tyr Glu Ala Val Ser
                 35                  40                  45
Glu Thr Met Gln Ser Ala Thr Gly Ile Gln Tyr Ser Val Thr Pro
                 50                  55                  60
Ser Tyr Gln Val Ser Ala Met Pro Gln Ser Ser Gly Ser His Gly
```

```
                    65                  70                  75
Pro Ala Ile Ala Ala Val His Ser Ser His His His Pro Thr Ala
                    80                  85                  90
Val Gln Pro His Gly Gly Gln Val Val Gln Ser His Ala His Pro
                    95                 100                 105
Ala Pro Pro Val Ala Pro Val Gln Gly Gln Gln Gln Phe Gln Arg
                   110                 115                 120
Leu Lys Val Glu Asp Ala Leu Ser Tyr Leu Asp Gln Val Lys Leu
                   125                 130                 135
Gln Phe Gly Ser Gln Pro Gln Val Tyr Asn Asp Phe Leu Asp Ile
                   140                 145                 150
Met Lys Glu Phe Lys Ser Gln Ser Ile Asp Thr Pro Gly Val Ile
                   155                 160                 165
Ser Arg Val Ser Gln Leu Phe Lys Gly His Pro Asp Leu Ile Met
                   170                 175                 180
Gly Phe Asn Thr Phe Leu Pro Pro Gly Tyr Lys Ile Glu Val Gln
                   185                 190                 195
Thr Asn Asp Met Val Asn Val Thr Thr Pro Gly Gln Val His Gln
                   200                 205                 210
Ile Pro Thr His Gly Ile Gln Pro Gln Pro Gln Pro Pro Pro Gln
                   215                 220                 225
His Pro Ser Gln Pro Ser Ala Gln Ser Ala Pro Ala Pro Ala Gln
                   230                 235                 240
Pro Ala Pro Gln Pro Pro Pro Ala Lys Val Ser Lys Pro Ser Gln
                   245                 250                 255
Leu Gln Ala His Thr Pro Ala Ser Gln Gln Thr Pro Pro Leu Pro
                   260                 265                 270
Pro Tyr Ala Ser Pro Arg Ser Pro Pro Val Gln Pro His Thr Pro
                   275                 280                 285
Val Thr Ile Ser Leu Gly Thr Ala Pro Ser Leu Gln Asn Asn Gln
                   290                 295                 300
Pro Val Glu Phe Asn His Ala Ile Asn Tyr Val Asn Lys Ile Lys
                   305                 310                 315
Asn Arg Phe Gln Gly Gln Pro Asp Ile Tyr Lys Ala Phe Leu Glu
                   320                 325                 330
Ile Leu His Thr Tyr Gln Lys Glu Gln Arg Asn Ala Lys Glu Ala
                   335                 340                 345
Gly Gly Asn Tyr Thr Pro Ala Leu Thr Glu Gln Glu Val Tyr Ala
                   350                 355                 360
Gln Val Ala Arg Leu Phe Lys Asn Gln Glu Asp Leu Leu Ser Glu
                   365                 370                 375
Phe Gly Gln Phe Leu Pro Asp Ala Asn Ser Ser Val Leu Leu Ser
                   380                 385                 390
Lys Thr Thr Ala Glu Lys Val Asp Ser Val Arg Asn Asp His Gly
                   395                 400                 405
Gly Thr Val Lys Lys Pro Gln Leu Asn Asn Lys Pro Gln Arg Pro
                   410                 415                 420
Ser Gln Asn Gly Cys Gln Ile Arg Arg His Pro Thr Gly Thr Thr
                   425                 430                 435
Pro Pro Val Lys Lys Lys Pro Lys Leu Leu Asn Leu Lys Asp Ser
                   440                 445                 450
Ser Met Ala Asp Ala Ser Lys His Gly Gly Gly Thr Glu Ser Leu
                   455                 460                 465
Phe Phe Asp Lys Val Arg Lys Ala Leu Arg Ser Ala Glu Ala Tyr
                   470                 475                 480
Glu Asn Phe Leu Arg Cys Leu Val Ile Phe Asn Gln Glu Val Ile
                   485                 490                 495
Ser Arg Ala Glu Leu Val Gln Leu Val Ser Pro Phe Leu Gly Lys
                   500                 505                 510
Phe Pro Glu Leu Phe Asn Trp Phe Lys Asn Phe Leu Gly Tyr Lys
                   515                 520                 525
Glu Ser Val His Leu Glu Thr Tyr Pro Lys Glu Arg Ala Thr Glu
                   530                 535                 540
```

```
Gly Ile Ala Met Glu Ile Asp Tyr Ala Ser Cys Lys Arg Leu Gly
                545                 550                 555
Ser Ser Tyr Arg Ala Leu Pro Lys Ser Tyr Gln Gln Pro Lys Cys
                560                 565                 570
Thr Gly Arg Thr Pro Leu Cys Lys Glu Val Leu Asn Asp Thr Trp
                575                 580                 585
Val Ser Phe Pro Ser Trp Ser Glu Asp Ser Thr Phe Val Ser Ser
                590                 595                 600
Lys Lys Thr Gln Tyr Glu Glu His Ile Tyr Arg Cys Glu Asp Glu
                605                 610                 615
Arg Phe Glu Leu Asp Val Val Leu Glu Thr Asn Leu Ala Thr Ile
                620                 625                 630
Arg Val Leu Glu Ala Ile Gln Lys Lys Leu Ser Arg Leu Ser Ala
                635                 640                 645
Glu Glu Gln Ala Lys Phe Arg Leu Asp Asn Thr Leu Gly Gly Thr
                650                 655                 660
Ser Glu Val Ile His Arg Lys Ala Leu Gln Arg Ile Tyr Ala Asp
                665                 670                 675
Lys Ala Ala Asp Ile Ile Asp Gly Leu Arg Lys Asn Pro Ser Ile
                680                 685                 690
Ala Val Pro Ile Val Leu Lys Arg Leu Lys Met Lys Glu Glu Glu
                695                 700                 705
Trp Arg Glu Ala Gln Arg Gly Phe Asn Lys Val Trp Arg Glu Gln
                710                 715                 720
Asn Glu Lys Tyr Tyr Leu Lys Ser Leu Asp His Gln Gly Ile Asn
                725                 730                 735
Phe Lys Gln Asn Asp Thr Lys Val Leu Arg Ser Lys Ser Leu Leu
                740                 745                 750
Asn Glu Ile Glu Ser Ile Tyr Asp Glu Arg Gln Glu Gln Ala Thr
                755                 760                 765
Glu Glu Asn Ala Gly Val Pro Val Gly Pro His Leu Ser Leu Ala
                770                 775                 780
Tyr Glu Asp Lys Gln Ile Leu Glu Asp Ala Ala Ala Leu Ile Ile
                785                 790                 795
His His Val Lys Arg Gln Thr Gly Ile Gln Lys Glu Asp Lys Tyr
                800                 805                 810
Lys Ile Lys Gln Ile Met His His Phe Ile Pro Asp Leu Leu Phe
                815                 820                 825
Ala Gln Arg Gly Asp Leu Ser Asp Val Glu Glu Glu Glu Glu Glu
                830                 835                 840
Glu Met Asp Val Asp Glu Ala Thr Gly Ala Val Lys Lys His Asn
                845                 850                 855
Gly Val Gly Gly Ser Pro Pro Lys Ser Lys Leu Leu Phe Ser Asn
                860                 865                 870
Thr Ala Ala Gln Lys Leu Arg Gly Met Asp Glu Val Tyr Asn Leu
                875                 880                 885
Phe Tyr Val Asn Asn Asn Trp Tyr Ile Phe Met Arg Leu His Gln
                890                 895                 900
Ile Leu Cys Leu Arg Leu Leu Arg Ile Cys Ser Gln Ala Glu Arg
                905                 910                 915
Gln Ile Glu Glu Glu Asn Arg Glu Arg Glu Trp Glu Arg Glu Val
                920                 925                 930
Leu Gly Ile Lys Arg Asp Lys Ser Asp Ser Pro Ala Ile Gln Leu
                935                 940                 945
Arg Leu Lys Glu Pro Met Asp Val Asp Val Glu Asp Tyr Tyr Pro
                950                 955                 960
Ala Phe Leu Asp Met Val Arg Ser Leu Leu Asp Gly Asn Ile Asp
                965                 970                 975
Ser Ser Gln Tyr Glu Asp Ser Leu Arg Glu Met Phe Thr Ile His
                980                 985                 990
Ala Tyr Ile Ala Phe Thr Met Asp Lys Leu Ile Gln Ser Ile Val
                995                1000                1005
Arg Gln Leu Gln His Ile Val Ser Asp Glu Ile Cys Val Gln Val
```

```
                1010                1015                1020
Thr Asp Leu Tyr Leu Ala Glu Asn Asn Asn Gly Ala Thr Gly Gly
                1025                1030                1035
Gln Leu Asn Thr Gln Asn Ser Arg Ser Leu Leu Glu Ser Thr Tyr
                1040                1045                1050
Gln Arg Lys Ala Glu Gln Leu Met Ser Asp Glu Asn Cys Phe Lys
                1055                1060                1065
Leu Met Phe Ile Gln Ser Gln Gly Gln Val Gln Leu Thr Ile Glu
                1070                1075                1080
Leu Leu Asp Thr Glu Glu Glu Asn Ser Asp Asp Pro Val Glu Ala
                1085                1090                1095
Glu Arg Trp Ser Asp Tyr Val Glu Arg Tyr Met Asn Ser Asp Thr
                1100                1105                1110
Thr Ser Pro Glu Leu Arg Glu His Leu Ala Gln Lys Pro Val Phe
                1115                1120                1125
Leu Pro Arg Asn Leu Arg Arg Ile Arg Lys Cys Gln Arg Gly Arg
                1130                1135                1140
Glu Gln Gln Glu Lys Glu Gly Lys Glu Gly Asn Ser Lys Lys Thr
                1145                1150                1155
Met Glu Asn Val Asp Ser Leu Asp Lys Leu Glu Cys Arg Phe Lys
                1160                1165                1170
Leu Asn Ser Tyr Lys Met Val Tyr Val Ile Lys Ser Glu Asp Tyr
                1175                1180                1185
Met Tyr Arg Arg Thr Ala Leu Leu Arg Ala His Gln Ser His Glu
                1190                1195                1200
Arg Val Ser Lys Arg Leu His Gln Arg Phe Gln Ala Trp Val Asp
                1205                1210                1215
Lys Trp Thr Lys Glu His Val Pro Arg Glu Met Ala Ala Glu Thr
                1220                1225                1230
Ser Lys Trp Leu Met Gly Glu Gly Leu Glu Gly Leu Val Pro Cys
                1235                1240                1245
Thr Thr Thr Cys Asp Thr Glu Thr Leu His Phe Val Ser Ile Asn
                1250                1255                1260
Lys Tyr Arg Val Lys Tyr Gly Thr Val Phe Lys Ala Pro
                1265                1270
```

<210> 9
<211> 381
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4019390CD1

<400> 9

```
Met Glu Pro Leu Thr Phe Lys Asp Val Ala Ile Glu Phe Ser Leu
  1               5                  10                  15
Glu Glu Trp Gln Cys Leu Asp Thr Ala Gln Arg Asp Leu Tyr Arg
                20                  25                  30
Asn Val Leu Leu Glu Asn Tyr Arg Asn Leu Val Phe Leu Gly Ile
                35                  40                  45
Ala Val Ser Lys Pro Tyr Leu Ile Thr Cys Leu Glu Gln Lys Lys
                50                  55                  60
Glu Pro Trp Asn Ile Lys Arg His Glu Met Val Ala Lys Pro Pro
                65                  70                  75
Val Met Ser Phe His Phe Ala Gln Asp Leu Trp Pro Glu Gln Asn
                80                  85                  90
Ile Lys Asp Ser Phe Gln Lys Val Thr Leu Arg Arg Tyr Gly Lys
                95                 100                 105
Cys Glu Tyr Glu Asn Leu Gln Leu Arg Lys Gly Cys Lys His Val
               110                 115                 120
Asp Glu Cys Thr Gly His Lys Gly Gly His Asn Thr Val Asn Gln
```

```
                125                 130                 135
Cys Leu Thr Ala Thr Pro Ser Lys Ile Phe Gln Cys Asn Lys Tyr
                140                 145                 150
Val Lys Val Phe Asp Lys Phe Ser Asn Ser Asn Arg Tyr Lys Arg
                155                 160                 165
Arg His Thr Gly Asn Lys His Phe Lys Cys Lys Glu Cys Ser Lys
                170                 175                 180
Ser Phe Cys Val Leu Ser Gln Leu Thr Gln His Arg Arg Ile His
                185                 190                 195
Thr Arg Val Asn Ser Tyr Lys Cys Glu Glu Cys Gly Lys Ala Phe
                200                 205                 210
Asn Trp Phe Ser Thr Leu Thr Lys His Lys Arg Ile His Thr Gly
                215                 220                 225
Glu Lys Pro Tyr Lys Cys Glu Glu Cys Gly Lys Ala Phe Asn Gln
                230                 235                 240
Ser Ser Gln Leu Thr Arg His Lys Ile Ile His Thr Glu Glu Lys
                245                 250                 255
Pro Asn Lys Cys Glu Glu Cys Gly Lys Ala Phe Lys Gln Ala Ser
                260                 265                 270
His Leu Thr Ile His Lys Ile Ile His Thr Gly Glu Lys Pro Tyr
                275                 280                 285
Lys Tyr Glu Glu Cys Gly Lys Val Phe Ser Gln Ser Ser His Leu
                290                 295                 300
Thr Thr Gln Lys Ile Leu His Thr Gly Glu Asn Leu Tyr Lys Cys
                305                 310                 315
Lys Glu Cys Gly Lys Ala Phe Asn Leu Phe Ser Asn Leu Thr Asn
                320                 325                 330
His Lys Arg Ile His Ala Gly Glu Lys Pro Tyr Lys Cys Lys Glu
                335                 340                 345
Cys Gly Arg Ala Phe Asn Ile Ser Ser Asn Leu Asn Lys Gln Glu
                350                 355                 360
Lys Ile His Thr Gly Gly Lys Leu Asn Lys Cys Glu Glu Cys Asp
                365                 370                 375
Lys Leu Leu Thr Asp Pro
                380

<210> 10
<211> 290
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 986452CD1

<400> 10
Met Ser Ser Glu Ala Glu Thr Gln Gln Pro Pro Ala Ala Pro Ala
  1                 5                  10                  15
Leu Ser Ala Ala Asp Thr Lys Pro Gly Thr Thr Gly Ser Gly Ala
                 20                  25                  30
Gly Ser Gly Gly Pro Gly Gly Leu Thr Ser Ala Ala Pro Ala Gly
                 35                  40                  45
Gly Asp Lys Lys Val Ile Ala Thr Lys Val Leu Gly Thr Val Lys
                 50                  55                  60
Trp Phe Asn Val Arg Asn Gly Tyr Gly Phe Ile Asn Arg Asn Asp
                 65                  70                  75
Thr Lys Glu Asp Val Phe Val His Gln Gly Ala Glu Ala Ala Asn
                 80                  85                  90
Val Thr Gly Pro Gly Gly Val Pro Val Gln Gly Ser Lys Tyr Ala
                 95                 100                 105
Ala Asp Arg Asn His Tyr Arg Arg Tyr Pro Arg Arg Arg Gly Pro
                110                 115                 120
Pro Arg Asn Tyr Gln Gln Asn Tyr Gln Asn Ser Glu Ser Gly Glu
```

```
                  125                 130                 135
Lys Asn Glu Gly Ser Glu Ser Ala Pro Glu Gly Gln Ala Gln Gln
                  140                 145                 150
Arg Arg Pro Tyr Arg Arg Arg Arg Phe Pro Pro Tyr Tyr Met Arg
                  155                 160                 165
Arg Pro Tyr Gly Arg Arg Pro Gln Tyr Ser Asn Pro Pro Val Gln
                  170                 175                 180
Gly Glu Val Met Glu Gly Ala Asp Asn Gln Gly Ala Gly Glu Gln
                  185                 190                 195
Gly Arg Pro Val Arg Gln Asn Met Tyr Arg Gly Tyr Arg Pro Arg
                  200                 205                 210
Phe Arg Arg Gly Pro Pro Arg Gln Arg Gln Pro Arg Glu Asp Gly
                  215                 220                 225
Asn Glu Glu Asp Lys Glu Asn Gln Gly Asp Glu Thr Gln Gly Gln
                  230                 235                 240
Gln Pro Pro Gln Arg Arg Tyr Arg Arg Asn Phe Asn Tyr Arg Arg
                  245                 250                 255
Arg Arg Pro Glu Asn Pro Lys Pro Gln Asp Gly Lys Glu Thr Lys
                  260                 265                 270
Ala Ala Asp Pro Pro Ala Glu Asn Ser Ser Ala Pro Glu Ala Glu
                  275                 280                 285
Gln Gly Gly Ala Glu
                  290

<210> 11
<211> 588
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2807579CD1

<400> 11
Met Gly Leu Tyr Gly Gln Ala Cys Pro Ser Val Thr Ser Leu Arg
  1                 5                  10                  15
Met Thr Ser Glu Leu Glu Ser Ser Leu Thr Ser Met Asp Trp Leu
                 20                  25                  30
Pro Gln Leu Thr Met Arg Ala Ala Ile Gln Lys Ser Asp Ala Thr
                 35                  40                  45
Gln Asn Ala His Gly Thr Gly Ile Ser Lys Lys Asn Ala Leu Leu
                 50                  55                  60
Asp Pro Asn Thr Thr Leu Asp Gln Glu Glu Val Gln Gln His Lys
                 65                  70                  75
Asp Gly Lys Pro Pro Tyr Ser Tyr Ala Ser Leu Ile Thr Phe Ala
                 80                  85                  90
Ile Asn Ser Ser Pro Lys Lys Lys Met Thr Leu Ser Glu Ile Tyr
                 95                 100                 105
Gln Trp Ile Cys Asp Asn Phe Pro Tyr Tyr Arg Glu Ala Gly Ser
                110                 115                 120
Gly Trp Lys Asn Ser Ile Arg His Asn Leu Ser Leu Asn Lys Cys
                125                 130                 135
Phe Leu Lys Val Pro Arg Ser Lys Asp Asp Pro Gly Lys Gly Ser
                140                 145                 150
Tyr Trp Ala Ile Asp Thr Asn Pro Lys Glu Asp Ala Leu Pro Thr
                155                 160                 165
Arg Pro Lys Lys Arg Ala Arg Ser Val Glu Arg Val Thr Leu Tyr
                170                 175                 180
Asn Thr Asp Gln Asp Gly Ser Asp Ser Pro Arg Ser Ser Leu Asn
                185                 190                 195
Asn Ser Leu Ser Asp Gln Ser Leu Ala Ser Val Asn Leu Asn Ser
                200                 205                 210
Val Gly Ser Val His Ser Tyr Thr Pro Val Thr Ser His Pro Glu
```

```
                215                 220                 225
Ser Val Ser Gln Ser Leu Thr Pro Gln Gln Gln Pro Gln Tyr Asn
                230                 235                 240
Leu Pro Glu Arg Asp Lys Gln Leu Leu Phe Ser Glu Tyr Asn Phe
                245                 250                 255
Glu Asp Leu Ser Ala Ser Phe Arg Ser Leu Tyr Lys Ser Val Phe
                260                 265                 270
Glu Gln Ser Leu Ser Gln Gln Gly Leu Met Asn Ile Pro Ser Glu
                275                 280                 285
Ser Ser Gln Gln Ser His Thr Ser Cys Thr Tyr Gln His Ser Pro
                290                 295                 300
Ser Ser Thr Val Ser Thr His Pro His Ser Asn Gln Ser Ser Leu
                305                 310                 315
Ser Asn Ser His Gly Ser Gly Leu Asn Thr Thr Gly Ser Asn Ser
                320                 325                 330
Val Ala Gln Val Ser Leu Ser His Pro Gln Met His Thr Gln Pro
                335                 340                 345
Ser Pro His Pro Pro His Arg Pro His Gly Leu Pro Gln His Pro
                350                 355                 360
Gln Arg Ser Pro His Pro Ala Pro His Pro Gln Gln His Ser Gln
                365                 370                 375
Leu Gln Ser Pro His Pro Gln His Pro Ser Pro His Gln His Ile
                380                 385                 390
Gln His His Pro Asn His Gln His Gln Thr Leu Thr His Gln Ala
                395                 400                 405
Pro Pro Pro Pro Gln Gln Val Ser Cys Asn Ser Gly Val Ser Asn
                410                 415                 420
Asp Trp Tyr Ala Thr Leu Asp Met Leu Lys Glu Ser Cys Arg Ile
                425                 430                 435
Ala Ser Ser Val Asn Trp Ser Asp Val Asp Leu Ser Gln Phe Gln
                440                 445                 450
Gly Leu Met Glu Ser Met Arg Gln Ala Asp Leu Lys Asn Trp Ser
                455                 460                 465
Leu Asp Gln Val Gln Phe Ala Asp Leu Cys Ser Ser Leu Asn Gln
                470                 475                 480
Phe Phe Thr Gln Thr Gly Leu Ile His Ser Gln Ser Asn Val Gln
                485                 490                 495
Gln Asn Val Cys His Gly Ala Met His Pro Thr Lys Pro Ser Gln
                500                 505                 510
His Ile Gly Thr Gly Asn Leu Tyr Ile Asp Ser Arg Gln Asn Leu
                515                 520                 525
Pro Pro Ser Val Met Pro Pro Pro Gly Tyr Pro His Ile Pro Gln
                530                 535                 540
Ala Leu Ser Thr Pro Gly Thr Thr Met Ala Gly His His Arg Ala
                545                 550                 555
Met Asn Gln Gln His Met Met Pro Ser Gln Ala Phe Gln Met Arg
                560                 565                 570
Arg Ser Leu Pro Pro Asp Asp Ile Gln Asp Asp Phe Asp Trp Asp
                575                 580                 585
Ser Ile Val
```

<210> 12
<211> 103
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5724273CD1

<400> 12
Met Cys Leu Ile Thr Leu Asn Asn Ser Tyr Arg Met Phe Glu Ser

```
  1               5                  10                  15
Cys Ser Gly Phe Ser Gly Phe Cys Ala Ser Pro Ile Glu Glu Ser
                20                  25                  30
His Gly Ala Leu Ile Ser Ser Cys Asn Ser Arg Thr Met Thr Asp
                35                  40                  45
Gly Leu Val Thr Phe Arg Asp Val Ala Ile Asp Phe Ser Gln Glu
                50                  55                  60
Glu Trp Glu Cys Leu Asp Pro Ala Gln Arg Asp Leu Tyr Val Asp
                65                  70                  75
Val Met Leu Glu Asn Tyr Ser Asn Leu Val Ser Leu Asp Leu Glu
                80                  85                  90
Ser Lys Thr Tyr Glu Thr Lys Lys Asn Ile Phe Arg Lys
                95                 100
```

<210> 13
<211> 593
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3614884CD1

<400> 13

```
Met Thr Met Phe Lys Glu Ala Val Thr Phe Lys Asp Val Ala Val
  1               5                  10                  15
Val Phe Thr Glu Glu Glu Leu Gly Leu Leu Asp Val Ser Gln Arg
                20                  25                  30
Lys Leu Tyr Arg Asp Val Met Leu Glu Asn Phe Arg Asn Leu Leu
                35                  40                  45
Ser Val Gly His Gln Leu Ser His Arg Asp Thr Phe His Phe Gln
                50                  55                  60
Arg Glu Glu Lys Phe Trp Ile Met Glu Thr Ala Thr Gln Arg Glu
                65                  70                  75
Gly Asn Ser Gly Gly Lys Ile Gln Thr Glu Leu Glu Ser Val Pro
                80                  85                  90
Glu Thr Gly Pro His Glu Glu Trp Ser Cys Gln Gln Ile Trp Glu
                95                 100                 105
Gln Thr Ala Ser Glu Leu Thr Arg Pro Gln Asp Ser Ile Ser Ser
               110                 115                 120
Ser Gln Phe Ser Thr Gln Gly Asp Val Pro Ser Gln Val Asp Ala
               125                 130                 135
Gly Leu Ser Ile Ile His Ile Gly Glu Thr Pro Ser Glu His Gly
               140                 145                 150
Lys Cys Lys Lys Phe Phe Ser Asp Val Ser Ile Leu Asp Leu His
               155                 160                 165
Gln Gln Leu His Ser Gly Lys Ile Ser His Thr Cys Asn Glu Tyr
               170                 175                 180
Arg Lys Arg Phe Cys Tyr Ser Ser Ala Leu Cys Leu His Gln Lys
               185                 190                 195
Val His Met Gly Glu Lys Arg Tyr Lys Cys Asp Val Cys Ser Lys
               200                 205                 210
Ala Phe Ser Gln Asn Ser Gln Leu Gln Thr His Gln Arg Ile His
               215                 220                 225
Thr Gly Glu Lys Pro Phe Lys Cys Glu Gln Cys Gly Lys Ser Phe
               230                 235                 240
Ser Arg Arg Ser Gly Met Tyr Val His Cys Lys Leu His Thr Gly
               245                 250                 255
Glu Lys Pro His Ile Cys Glu Glu Cys Gly Lys Ala Phe Ile His
               260                 265                 270
Asn Ser Gln Leu Arg Glu His Gln Arg Ile His Thr Gly Glu Lys
               275                 280                 285
Pro Phe Lys Cys Tyr Ile Cys Gly Lys Ser Phe His Ser Arg Ser
```

```
                290                 295                 300
Asn Leu Asn Arg His Ser Met Val His Met Gln Glu Lys Ser Phe
                305                 310                 315
Arg Cys Asp Thr Cys Ser Asn Ser Phe Gly Gln Arg Ser Ala Leu
                320                 325                 330
Asn Ser His Cys Met Asp His Thr Lys Glu Lys Leu Tyr Lys Cys
                335                 340                 345
Glu Glu Cys Gly Arg Ser Phe Thr Cys Arg Gln Asp Leu Cys Lys
                350                 355                 360
His Gln Met Asp His Thr Gly Asp Lys Pro Tyr Asn Cys Asn Val
                365                 370                 375
Cys Gly Lys Gly Phe Arg Trp Ser Ser Cys Leu Ser Arg His Gln
                380                 385                 390
Arg Val His Asn Gly Glu Thr Thr Phe Lys Cys Asp Gly Cys Gly
                395                 400                 405
Lys Arg Phe Tyr Met Asn Ser Gln Gly His Ser His Gln Arg Ala
                410                 415                 420
Tyr Arg Glu Glu Glu Leu Tyr Lys Cys Gln Lys Cys Gly Lys Gly
                425                 430                 435
Tyr Ile Ser Lys Phe Asn Leu Asp Leu His Gln Arg Val His Thr
                440                 445                 450
Gly Glu Arg Pro Tyr Asn Cys Lys Glu Cys Gly Lys Ser Phe Arg
                455                 460                 465
Trp Ala Ser Gly Ile Leu Arg His Lys Arg Leu His Thr Gly Glu
                470                 475                 480
Lys Pro Phe Lys Cys Glu Glu Cys Gly Lys Arg Phe Thr Glu Asn
                485                 490                 495
Ser Lys Leu Arg Phe His Gln Arg Ile His Thr Gly Glu Lys Pro
                500                 505                 510
Tyr Lys Cys Glu Glu Cys Gly Lys Gly Phe Arg Trp Ala Ser Thr
                515                 520                 525
His Leu Thr His Gln Arg Leu His Ser Arg Glu Lys Leu Phe Gln
                530                 535                 540
Cys Glu Asp Cys Gly Lys Ser Ser Glu His Ser Ser Cys Leu Gln
                545                 550                 555
Asp Gln Gln Ser Asp His Ser Gly Glu Lys Thr Ser Lys Cys Glu
                560                 565                 570
Asp Cys Gly Lys Arg Tyr Glu Arg Arg Leu Asn Leu Asp Met Ile
                575                 580                 585
Leu Ser Leu Phe Leu Asn Asp Ile
                590
```

<210> 14
<211> 281
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3794954CD1

<400> 14

```
Met Leu Ser Gln Leu Glu Gly Gly Glu Glu Gln Trp Val Pro Asp
  1                 5                  10                  15
Pro Gln Asp Leu Glu Glu Arg Asp Ile Leu Arg Val Thr Tyr Thr
                 20                  25                  30
Gly Asp Gly Ser Glu His Glu Gly Asp Thr Pro Glu Leu Glu Ala
                 35                  40                  45
Glu Pro Pro Arg Met Leu Ser Ser Val Ser Glu Asp Thr Val Leu
                 50                  55                  60
Trp Asn Pro Glu His Asp Glu Ser Trp Asp Ser Met Pro Ser Ser
                 65                  70                  75
Ser Arg Gly Met Leu Leu Gly Pro Pro Phe Leu Gln Glu Asp Ser
```

```
                  80                  85                  90
Phe Ser Asn Leu Leu Cys Ser Thr Glu Met Asp Ser Leu Leu Arg
                  95                 100                 105
Pro His Thr Cys Pro Gln Cys Gly Lys Gln Phe Val Trp Gly Ser
                 110                 115                 120
His Leu Ala Arg His Gln Gln Thr His Thr Gly Glu Arg Pro Tyr
                 125                 130                 135
Ser Cys Leu Lys Cys Glu Lys Thr Phe Gly Arg Arg His His Leu
                 140                 145                 150
Ile Arg His Gln Lys Thr His Leu His Asp Lys Thr Ser Arg Cys
                 155                 160                 165
Ser Glu Cys Gly Lys Asn Phe Arg Cys Asn Ser His Leu Ala Ser
                 170                 175                 180
His Gln Arg Val His Ala Glu Gly Lys Ser Cys Lys Gly Gln Glu
                 185                 190                 195
Val Glu Arg Ala Leu Ala Gln Gly Asn Gly Arg Val Pro His Gln
                 200                 205                 210
Cys Gln Ser Val Thr Cys Ala Leu Asn Val Gly Lys Ser Phe Gly
                 215                 220                 225
Arg Arg His His Leu Val Arg His Trp Leu Thr His Thr Gly Glu
                 230                 235                 240
Lys Pro Phe Gln Cys Pro Arg Cys Glu Lys Ser Phe Gly Arg Lys
                 245                 250                 255
His His Leu Asp Arg His Leu Leu Thr His Gln Gly Gln Ser Pro
                 260                 265                 270
Arg Asn Ser Trp Asp Arg Gly Thr Ser Val Phe
                 275                 280
```

<210> 15
<211> 539
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7399016CD1

<400> 15

```
Met Gly His Cys Arg Leu Cys His Gly Lys Phe Ser Ser Arg Ser
  1               5                  10                  15
Leu Arg Ser Ile Ser Glu Arg Ala Pro Gly Ala Ser Met Glu Arg
                 20                  25                  30
Pro Ser Ala Glu Glu Arg Val Leu Val Arg Asp Phe Gln Arg Leu
                 35                  40                  45
Leu Gly Val Ala Val Arg Gln Asp Pro Thr Leu Ser Pro Phe Val
                 50                  55                  60
Cys Lys Ser Cys His Ala Gln Phe Tyr Gln Cys His Ser Leu Leu
                 65                  70                  75
Lys Ser Phe Leu Gln Arg Val Asn Ala Ser Pro Ala Gly Arg Arg
                 80                  85                  90
Lys Pro Cys Ala Lys Val Gly Ala Gln Pro Pro Thr Gly Ala Glu
                 95                 100                 105
Glu Gly Ala Cys Leu Val Asp Leu Ile Thr Ser Ser Pro Gln Cys
                110                 115                 120
Leu His Gly Leu Val Gly Trp Val His Gly His Ala Ala Ser Cys
                125                 130                 135
Gly Ala Leu Pro His Leu Gln Arg Thr Leu Ser Ser Glu Tyr Cys
                140                 145                 150
Gly Val Ile Gln Val Val Trp Gly Cys Asp Gln Gly His Asp Tyr
                155                 160                 165
Thr Met Asp Thr Ser Ser Ser Cys Lys Ala Phe Leu Leu Asp Ser
                170                 175                 180
Ala Leu Ala Val Lys Trp Pro Trp Asp Lys Glu Thr Ala Pro Arg
```

```
                185                 190                 195
Leu Pro Gln His Arg Gly Trp Asn Pro Gly Asp Ala Pro Gln Thr
                200                 205                 210
Ser Gln Gly Arg Gly Thr Gly Thr Pro Val Gly Ala Glu Thr Lys
                215                 220                 225
Thr Leu Pro Ser Thr Asp Val Ala Gln Pro Pro Ser Asp Ser Asp
                230                 235                 240
Ala Val Gly Pro Arg Ser Gly Phe Pro Pro Gln Pro Ser Leu Pro
                245                 250                 255
Leu Cys Arg Ala Pro Gly Gln Leu Gly Glu Lys Gln Leu Pro Ser
                260                 265                 270
Ser Thr Ser Asp Asp Arg Val Lys Asp Glu Phe Ser Asp Leu Ser
                275                 280                 285
Glu Gly Asp Val Leu Ser Glu Asp Glu Asn Asp Lys Lys Gln Asn
                290                 295                 300
Ala Gln Ser Ser Asp Glu Ser Phe Glu Pro Tyr Pro Glu Arg Lys
                305                 310                 315
Val Ser Gly Lys Lys Ser Glu Ser Lys Glu Ala Lys Lys Ser Glu
                320                 325                 330
Glu Pro Arg Ile Arg Lys Lys Pro Gly Pro Lys Pro Gly Trp Lys
                335                 340                 345
Lys Lys Leu Arg Cys Glu Arg Glu Glu Leu Pro Thr Ile Tyr Lys
                350                 355                 360
Cys Pro Tyr Gln Gly Cys Thr Ala Val Tyr Arg Gly Ala Asp Gly
                365                 370                 375
Met Lys Lys His Ile Lys Glu His His Glu Glu Val Arg Glu Arg
                380                 385                 390
Pro Cys Pro His Pro Gly Cys Asn Lys Val Phe Met Ile Asp Arg
                395                 400                 405
Tyr Leu Gln Arg His Val Lys Leu Ile His Thr Glu Val Arg Asn
                410                 415                 420
Tyr Ile Cys Asp Glu Cys Gly Gln Thr Phe Lys Gln Arg Lys His
                425                 430                 435
Leu Leu Val His Gln Met Arg His Ser Gly Ala Lys Pro Leu Gln
                440                 445                 450
Cys Glu Val Cys Gly Phe Gln Cys Arg Gln Arg Ala Ser Leu Lys
                455                 460                 465
Tyr His Met Thr Lys His Lys Ala Glu Thr Glu Leu Asp Phe Ala
                470                 475                 480
Cys Asp Gln Cys Gly Arg Arg Phe Glu Lys Ala His Asn Leu Asn
                485                 490                 495
Val His Met Ser Met Val His Pro Leu Thr Gln Thr Gln Asp Lys
                500                 505                 510
Ala Leu Pro Leu Glu Ala Glu Pro Pro Pro Gly Pro Pro Ser Pro
                515                 520                 525
Ser Val Thr Thr Glu Gly Gln Ala Val Lys Pro Glu Pro Thr
                530                 535

<210> 16
<211> 390
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6996690CD1

<400> 16
Met Ala Glu Ile His Asn Gly Gly Glu Leu Cys Asp Phe Met Glu
  1               5                  10                  15
Asn Gly Glu Ile Phe Ser Glu His Ser Cys Leu Asn Ala His Met
                 20                  25                  30
Gly Thr Glu Asn Thr Gly Asp Thr Tyr Asp Cys Asp Glu Tyr Gly
```

```
                35                  40                  45
Glu Asn Phe Pro Met Leu His Asn Ser Ala Pro Ala Gly Glu Thr
                50                  55                  60
Leu Ser Val Leu Asn Gln Cys Arg Lys Ala Phe Ser Leu Pro Pro
                65                  70                  75
Asn Val His Gln Arg Thr Trp Ile Gly Asp Lys Ser Phe Glu Tyr
                80                  85                  90
Ser Asp Cys Glu Glu Ala Phe Val Asp Gln Ser His Leu Gln Ala
                95                 100                 105
Asn Arg Ile Thr His Asn Gly Glu Thr Leu Tyr Glu Gln Lys Gln
               110                 115                 120
Cys Gly Arg Ala Phe Thr Tyr Ser Thr Ser His Ala Val Ser Val
               125                 130                 135
Lys Met His Thr Val Glu Lys Pro Tyr Glu Cys Lys Glu Cys Gly
               140                 145                 150
Lys Phe Phe Arg Tyr Ser Ser Tyr Leu Asn Ser His Met Arg Thr
               155                 160                 165
His Thr Gly Glu Lys Pro Tyr Glu Cys Lys Glu Cys Gly Lys Cys
               170                 175                 180
Phe Thr Val Ser Ser His Leu Val Glu His Val Arg Ile His Thr
               185                 190                 195
Gly Glu Lys Pro Tyr Gln Cys Lys Glu Cys Gly Arg Ala Phe Ala
               200                 205                 210
Gly Arg Ser Gly Leu Thr Lys His Val Arg Ile His Thr Gly Glu
               215                 220                 225
Lys Pro Tyr Glu Cys Asn Glu Cys Gly Lys Ala Tyr Asn Arg Phe
               230                 235                 240
Tyr Leu Leu Thr Glu His Phe Lys Thr His Thr Glu Glu Lys Pro
               245                 250                 255
Phe Glu Cys Lys Val Cys Gly Lys Ser Phe Arg Ser Ser Ser Cys
               260                 265                 270
Leu Lys Asn His Phe Arg Ile His Thr Gly Ile Lys Pro Tyr Lys
               275                 280                 285
Cys Lys Glu Cys Gly Lys Ala Phe Thr Val Ser Ser Ser Leu His
               290                 295                 300
Asn His Val Lys Ile His Thr Gly Glu Lys Pro Tyr Glu Cys Lys
               305                 310                 315
Asp Cys Gly Lys Ala Phe Ala Thr Ser Ser Gln Leu Ile Glu His
               320                 325                 330
Ile Arg Thr His Thr Gly Glu Lys Pro Tyr Ile Cys Lys Glu Cys
               335                 340                 345
Gly Lys Thr Phe Arg Ala Ser Ser His Leu Gln Lys His Val Arg
               350                 355                 360
Ile His Thr Gly Glu Lys Pro Tyr Ile Cys Asn Glu Cys Gly Lys
               365                 370                 375
Ala Tyr Asn Arg Phe Tyr Leu Leu Thr Lys His Leu Lys Thr His
               380                 385                 390
```

<210> 17
<211> 807
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7740866CD1

<400> 17

```
Met Lys Glu Trp Lys Ser Lys Met Glu Ile Ser Glu Glu Lys Lys
  1                 5                  10                  15
Ser Ala Arg Ala Ala Ser Glu Lys Leu Gln Arg Gln Ile Thr Gln
                20                  25                  30
```

```
Glu Cys Glu Leu Val Glu Thr Ser Asn Ser Glu Asp Arg Leu Leu
                35                  40                  45
Lys His Trp Val Ser Pro Leu Lys Asp Ala Met Arg His Leu Pro
                50                  55                  60
Ser Gln Glu Ser Gly Ile Arg Glu Met His Ile Ile Pro Gln Lys
                65                  70                  75
Ala Ile Val Gly Glu Ile Gly His Gly Cys Asn Glu Gly Glu Lys
                80                  85                  90
Ile Leu Ser Ala Gly Glu Ser Ser His Arg Tyr Glu Val Ser Gly
                95                 100                 105
Gln Asn Phe Lys Gln Lys Ser Gly Leu Thr Glu His Gln Lys Ile
               110                 115                 120
His Asn Ile Asn Lys Thr Tyr Glu Cys Lys Glu Cys Glu Lys Thr
               125                 130                 135
Phe Asn Arg Ser Ser Asn Leu Ile Ile His Gln Arg Ile His Thr
               140                 145                 150
Gly Asn Lys Pro Tyr Val Cys Asn Glu Cys Gly Lys Asp Ser Asn
               155                 160                 165
Gln Ser Ser Asn Leu Ile Ile His Gln Arg Ile His Thr Gly Lys
               170                 175                 180
Lys Pro Tyr Ile Cys His Glu Cys Gly Lys Asp Phe Asn Gln Ser
               185                 190                 195
Ser Asn Leu Val Arg His Lys Gln Ile His Ser Gly Gly Asn Pro
               200                 205                 210
Tyr Glu Cys Lys Glu Cys Gly Lys Ala Phe Lys Gly Ser Ser Asn
               215                 220                 225
Leu Val Leu His Gln Arg Ile His Ser Arg Gly Lys Pro Tyr Leu
               230                 235                 240
Cys Asn Lys Cys Gly Lys Ala Phe Ser Gln Ser Thr Asp Leu Ile
               245                 250                 255
Ile His His Arg Ile His Thr Gly Glu Lys Pro Tyr Glu Cys Tyr
               260                 265                 270
Asp Cys Gly Gln Met Phe Ser Gln Ser Ser His Leu Val Pro His
               275                 280                 285
Gln Arg Ile His Thr Gly Glu Lys Pro Leu Lys Cys Asn Glu Cys
               290                 295                 300
Glu Lys Ala Phe Arg Gln His Ser His Leu Thr Glu His Gln Arg
               305                 310                 315
Leu His Ser Gly Glu Lys Pro Tyr Glu Cys His Arg Cys Gly Lys
               320                 325                 330
Thr Phe Ser Gly Arg Thr Ala Phe Leu Lys His Gln Arg Leu His
               335                 340                 345
Ala Gly Glu Lys Ile Glu Glu Cys Glu Lys Thr Phe Ser Lys Asp
               350                 355                 360
Glu Glu Leu Arg Glu Glu Gln Arg Ile His Gln Glu Glu Lys Ala
               365                 370                 375
Tyr Trp Cys Asn Gln Cys Gly Arg Asn Phe Gln Gly Thr Ser Asp
               380                 385                 390
Leu Ile Arg His Gln Val Thr His Thr Gly Glu Lys Pro Tyr Glu
               395                 400                 405
Cys Lys Glu Cys Gly Lys Thr Phe Asn Gln Ser Ser Asp Leu Leu
               410                 415                 420
Arg His His Arg Ile His Ser Gly Glu Lys Pro Cys Val Cys Ser
               425                 430                 435
Lys Cys Gly Lys Ser Phe Arg Gly Ser Ser Asp Leu Ile Arg His
               440                 445                 450
His Arg Val His Thr Gly Glu Lys Pro Tyr Glu Cys Ser Glu Cys
               455                 460                 465
Gly Lys Ala Phe Ser Gln Arg Ser His Leu Val Thr His Gln Lys
               470                 475                 480
Ile His Thr Gly Glu Lys Pro Tyr Gln Cys Thr Glu Cys Gly Lys
               485                 490                 495
Ala Phe Arg Arg Arg Ser Leu Leu Ile Gln His Arg Arg Ile His
```

```
                500                 505                 510
Ser Gly Glu Lys Pro Tyr Glu Cys Lys Glu Cys Gly Lys Leu Phe
                515                 520                 525
Ile Trp Arg Thr Ala Phe Leu Lys His Gln Ser Leu His Thr Gly
                530                 535                 540
Glu Lys Leu Glu Cys Glu Lys Thr Phe Ser Gln Asp Glu Glu Leu
                545                 550                 555
Arg Gly Glu Gln Lys Ile His Gln Glu Ala Lys Ala Tyr Trp Cys
                560                 565                 570
Asn Gln Cys Gly Arg Ala Phe Gln Gly Ser Ser Asp Leu Ile Arg
                575                 580                 585
His Gln Val Thr His Thr Arg Glu Lys Pro Tyr Glu Cys Lys Glu
                590                 595                 600
Cys Gly Lys Thr Phe Asn Gln Ser Ser Asp Leu Leu Arg His His
                605                 610                 615
Arg Ile His Ser Gly Glu Lys Pro Tyr Val Cys Asn Lys Cys Gly
                620                 625                 630
Lys Ser Phe Arg Gly Ser Ser Asp Leu Ile Lys His His Arg Ile
                635                 640                 645
His Thr Gly Glu Lys Pro Tyr Glu Cys Ser Glu Cys Gly Lys Ala
                650                 655                 660
Phe Ser Gln Arg Ser His Leu Ala Thr His Gln Lys Ile His Thr
                665                 670                 675
Gly Glu Lys Pro Tyr Gln Cys Ser Glu Cys Gly Asn Ala Phe Arg
                680                 685                 690
Arg Arg Ser Leu Leu Ile Gln His Arg Arg Leu His Ser Gly Glu
                695                 700                 705
Lys Pro Tyr Glu Cys Lys Glu Cys Gly Lys Leu Phe Met Trp His
                710                 715                 720
Thr Ala Phe Leu Lys His Gln Arg Leu His Ala Gly Glu Lys Leu
                725                 730                 735
Glu Glu Cys Glu Lys Thr Phe Ser Lys Asp Glu Glu Leu Arg Lys
                740                 745                 750
Glu Gln Arg Thr His Gln Glu Lys Lys Val Tyr Trp Cys Asn Gln
                755                 760                 765
Cys Ser Arg Thr Phe Gln Gly Ser Ser Asp Leu Ile Arg His Gln
                770                 775                 780
Val Thr His Thr Arg Glu Lys Pro Tyr Glu Cys Lys Glu Cys Gly
                785                 790                 795
Lys Thr Gln Ser Glu Leu Arg Pro Ser Glu Thr Ser
                800                 805

<210> 18
<211> 290
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 8181605CD1

<400> 18
Met Gly Glu Leu Ser Pro Ala Val Ala Gln Glu Glu Thr Pro Pro
  1                 5                  10                  15
Gly Asp Trp Leu Phe Gly Gly Val Arg Trp Gly Trp Asn Phe Arg
                 20                  25                  30
Cys Lys Pro Pro Val Gly Leu Asn Pro Arg Thr Gly Pro Glu Gly
                 35                  40                  45
Leu Pro Tyr Ser Ser Pro Asp Asn Gly Glu Ala Ile Leu Asp Pro
                 50                  55                  60
Ser Gln Ala Pro Arg Pro Phe Asn Glu Pro Cys Lys Tyr Pro Gly
                 65                  70                  75
Arg Thr Lys Gly Phe Gly His Lys Pro Gly Leu Lys Lys His Pro
```

```
                  80                  85                  90
Ala Ala Pro Pro Gly Gly Arg Pro Phe Thr Cys Ala Thr Cys Gly
                  95                 100                 105
Lys Ser Phe Gln Leu Gln Val Ser Leu Ser Ala His Gln Arg Ser
                 110                 115                 120
Cys Gly Ala Pro Asp Gly Ser Gly Pro Gly Thr Gly Gly Gly Gly
                 125                 130                 135
Ser Gly Ser Gly Gly Gly Gly Gly Gly Ser Gly Gly Gly Ser Ala
                 140                 145                 150
Arg Asp Gly Ser Ala Leu Arg Cys Gly Glu Cys Gly Arg Cys Phe
                 155                 160                 165
Thr Arg Pro Ala His Leu Ile Arg His Arg Met Leu His Thr Gly
                 170                 175                 180
Glu Arg Pro Phe Pro Cys Thr Glu Cys Glu Lys Arg Phe Thr Glu
                 185                 190                 195
Arg Ser Lys Leu Ile Asp His Tyr Arg Thr His Thr Gly Val Arg
                 200                 205                 210
Pro Phe Thr Cys Thr Val Cys Gly Lys Ser Phe Ile Arg Lys Asp
                 215                 220                 225
His Leu Arg Lys His Gln Arg Asn His Ala Ala Gly Ala Lys Thr
                 230                 235                 240
Pro Ala Arg Gly Gln Pro Leu Pro Thr Pro Pro Ala Pro Pro Asp
                 245                 250                 255
Pro Phe Lys Ser Pro Ala Ser Lys Gly Pro Leu Ala Ser Thr Asp
                 260                 265                 270
Leu Val Thr Asp Trp Thr Cys Gly Leu Ser Val Leu Gly Pro Thr
                 275                 280                 285
Asp Gly Gly Asp Met
                 290

<210> 19
<211> 452
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 8266487CD1

<400> 19
Met Lys Gly His Glu Gln Glu Ser Leu Phe Lys Cys Glu Val Cys
  1                 5                  10                  15
Ala Glu Arg Phe Pro Thr His Ala Lys Leu Ser Ser His Gln Arg
                  20                  25                  30
Ser His Phe Glu Pro Glu Arg Pro Tyr Lys Cys Asp Phe Pro Gly
                  35                  40                  45
Cys Glu Lys Thr Phe Ile Thr Val Ser Ala Leu Phe Ser His Asn
                  50                  55                  60
Arg Ala His Phe Arg Glu Gln Glu Leu Phe Ser Cys Ser Phe Pro
                  65                  70                  75
Gly Cys Ser Lys Gln Tyr Asp Lys Ala Cys Arg Leu Lys Ile His
                  80                  85                  90
Leu Arg Ser His Thr Gly Glu Arg Pro Phe Ile Cys Asp Ser Asp
                  95                 100                 105
Ser Cys Gly Trp Thr Phe Thr Ser Met Ser Lys Leu Leu Arg His
                 110                 115                 120
Arg Arg Lys His Asp Asp Asp Arg Arg Phe Thr Cys Pro Val Glu
                 125                 130                 135
Gly Cys Gly Lys Ser Phe Thr Arg Ala Glu His Leu Lys Gly His
                 140                 145                 150
Ser Ile Thr His Leu Gly Thr Lys Pro Phe Glu Cys Pro Val Glu
                 155                 160                 165
Gly Cys Cys Ala Arg Phe Ser Ala Arg Ser Ser Leu Tyr Ile His
```

```
                170                 175                 180
Ser Lys Lys His Val Gln Asp Val Gly Ala Pro Lys Ser Arg Cys
                185                 190                 195
Pro Val Ser Thr Cys Asn Arg Leu Phe Thr Ser Lys His Ser Met
                200                 205                 210
Lys Ala His Met Val Arg Gln His Ser Arg Arg Gln Asp Leu Leu
                215                 220                 225
Pro Gln Leu Glu Ala Pro Ser Ser Leu Thr Pro Ser Ser Glu Leu
                230                 235                 240
Ser Ser Pro Gly Gln Ser Glu Leu Thr Asn Met Asp Leu Ala Ala
                245                 250                 255
Leu Phe Ser Asp Thr Pro Ala Asn Ala Ser Gly Ser Ala Gly Gly
                260                 265                 270
Ser Asp Glu Ala Leu Asn Ser Gly Ile Leu Thr Ile Asp Val Thr
                275                 280                 285
Ser Val Ser Ser Ser Leu Gly Gly Asn Leu Pro Ala Asn Asn Ser
                290                 295                 300
Ser Leu Gly Pro Met Glu Pro Leu Val Leu Val Ala His Ser Asp
                305                 310                 315
Ile Pro Pro Ser Leu Asp Ser Pro Leu Val Leu Gly Thr Ala Ala
                320                 325                 330
Thr Val Leu Gln Gln Gly Ser Phe Ser Val Asp Asp Val Gln Thr
                335                 340                 345
Val Ser Ala Gly Ala Leu Gly Cys Leu Val Ala Leu Pro Met Lys
                350                 355                 360
Asn Leu Ser Asp Asp Pro Leu Ala Leu Thr Ser Asn Ser Asn Leu
                365                 370                 375
Ala Ala His Ile Thr Thr Pro Thr Ser Ser Ser Thr Pro Arg Glu
                380                 385                 390
Asn Ala Ser Val Pro Glu Leu Leu Ala Pro Ile Lys Val Glu Pro
                395                 400                 405
Asp Ser Pro Ser Arg Pro Gly Ala Val Gly Gln Gln Glu Gly Ser
                410                 415                 420
His Gly Leu Pro Gln Ser Thr Leu Pro Ser Pro Ala Glu Gln His
                425                 430                 435
Gly Ala Gln Asp Thr Glu Leu Ser Ala Gly Thr Gly Asn Phe Tyr
                440                 445                 450
Leu Val
```

<210> 20
<211> 259
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5552784CD1

```
<400> 20
Met Ala Ser Pro Gln Gly Gly Gln Ile Ala Ile Ala Met Arg Leu
  1                 5                  10                  15
Arg Asn Gln Leu Gln Ser Val Tyr Lys Met Asp Pro Leu Arg Asn
                 20                  25                  30
Glu Glu Glu Val Arg Val Lys Ile Lys Asp Leu Asn Glu His Ile
                 35                  40                  45
Val Cys Cys Leu Cys Ala Gly Tyr Phe Val Asp Ala Thr Thr Ile
                 50                  55                  60
Thr Glu Cys Leu His Thr Phe Cys Lys Ser Cys Ile Val Lys Tyr
                 65                  70                  75
Leu Gln Thr Ser Lys Tyr Cys Pro Met Cys Asn Ile Lys Ile His
                 80                  85                  90
Glu Thr Gln Pro Leu Leu Asn Leu Lys Leu Asp Arg Val Met Gln
```

```
                 95                  100                 105
Asp Ile Val Tyr Lys Leu Val Pro Gly Leu Gln Asp Ser Glu Glu
                110                 115                 120
Lys Arg Ile Arg Glu Phe Tyr Gln Ser Arg Gly Leu Asp Arg Val
                125                 130                 135
Thr Gln Pro Thr Gly Glu Glu Pro Ala Leu Ser Asn Leu Gly Leu
                140                 145                 150
Pro Phe Ser Ser Phe Asp His Ser Lys Ala His Tyr Tyr Arg Tyr
                155                 160                 165
Asp Glu Gln Leu Asn Leu Cys Leu Glu Arg Leu Ser Ser Gly Lys
                170                 175                 180
Asp Lys Asn Lys Ser Val Leu Gln Asn Lys Tyr Val Arg Cys Ser
                185                 190                 195
Val Arg Ala Glu Val Arg His Leu Arg Arg Val Leu Cys His Arg
                200                 205                 210
Leu Met Leu Asn Pro Gln His Val Gln Leu Leu Phe Asp Asn Glu
                215                 220                 225
Val Leu Pro Asp His Met Thr Met Lys Gln Ile Trp Leu Ser Arg
                230                 235                 240
Trp Phe Gly Lys Pro Ser Pro Leu Leu Leu Gln Tyr Ser Val Lys
                245                 250                 255
Glu Lys Arg Arg
```

<210> 21
<211> 665
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7281230CD1

<400> 21

```
Met Ala Ala Gln Met Ser Glu Ala Ser Ala Leu Ala Pro Gln Val
  1                   5                  10                  15
Phe Pro Ser Pro Leu Glu Leu Met Val Gly Glu Pro Ser Ser Lys
                 20                  25                  30
Ser Pro Gly Gln Cys Phe Trp Gly Phe Cys Tyr Glu Lys Ala Ala
                 35                  40                  45
Gly Pro Arg Gly Ala Leu Ala Gln Leu Arg Glu Leu Cys Cys Gln
                 50                  55                  60
Trp Leu Met Pro Glu Ala Cys Ser Lys Glu Gln Met Leu Glu Leu
                 65                  70                  75
Leu Val Leu Glu Gln Leu Leu Gly Thr Leu Leu Pro Glu Ile Gln
                 80                  85                  90
Ala Tyr Thr Gln Glu Gln Trp Leu Gly Ser Pro Glu Glu Ala Thr
                 95                 100                 105
Ala Leu Ala Glu Arg Leu Gln Gln Glu Ser Ala Gly Pro Gly Leu
                110                 115                 120
Gln Met Ser Gly Gly Trp Ser Gly Gly Trp Val Pro Ala Pro Arg
                125                 130                 135
Pro Gln Glu Glu Leu Val Pro Arg Thr Glu Glu Gly Glu Glu Gln
                140                 145                 150
Glu Ala Pro Leu Gly Pro Phe Gln Ala Pro Pro Pro Gly His Arg
                155                 160                 165
Arg Glu Met Glu Ser Pro Arg Gly Trp Thr Leu Gln Val Ala Pro
                170                 175                 180
Glu Glu Gly Gln Val Leu Cys Asn Val Lys Thr Ala Thr Arg Gly
                185                 190                 195
Leu Ser Glu Gly Ala Val Ser Gly Gly Trp Gly Ala Trp Glu Asn
                200                 205                 210
Ser Thr Glu Val Pro Arg Glu Ala Gly Asp Gly Gln Arg Gln Gln
```

```
                215                 220                 225
Ala Thr Leu Gly Ala Ala Asp Glu Gln Gly Gly Pro Gly Arg Glu
                230                 235                 240
Leu Gly Pro Arg Arg Arg Trp Ala Gly Arg Gly Trp Ala Gln Glu
                245                 250                 255
Arg Ala Cys Arg Pro Gly Val Ala Pro Phe Ala Ser Pro Gln Arg
                260                 265                 270
Ser Arg Ala Ala Gly Ala Gly Ser Ala Ala Arg Arg Ser Ala Arg
                275                 280                 285
Ala Leu Thr Cys Cys Ser Ser Ala Arg Ala Pro Gly Glu Lys Pro
                290                 295                 300
Tyr Thr Cys Pro Glu Cys Gly Lys Ala Phe Ala Trp Ser Ser Asn
                305                 310                 315
Leu Ser Gln His Gln Arg Ile His Ser Gly Glu Lys Pro Tyr Ala
                320                 325                 330
Cys Arg Glu Cys Gly Lys Ala Phe Arg Ala His Ser Gln Leu Ile
                335                 340                 345
His His Gln Glu Thr His Ser Gly Leu Lys Pro Phe Arg Cys Pro
                350                 355                 360
Asp Cys Gly Lys Ser Phe Gly Arg Ser Thr Thr Leu Val Gln His
                365                 370                 375
Arg Arg Thr His Thr Gly Glu Lys Pro Tyr Glu Cys Pro Glu Cys
                380                 385                 390
Gly Lys Ala Phe Ser Trp Asn Ser Asn Phe Leu Glu His Arg Arg
                395                 400                 405
Val His Thr Gly Ala Arg Pro His Ala Cys Arg Asp Cys Gly Lys
                410                 415                 420
Ala Phe Ser Gln Ser Ser Asn Leu Ala Glu His Leu Lys Ile His
                425                 430                 435
Ala Gly Ala Arg Pro His Ala Cys Pro Asp Cys Gly Lys Ala Phe
                440                 445                 450
Val Arg Val Ala Gly Leu Arg Gln His Arg Arg Thr His Ser Ser
                455                 460                 465
Glu Lys Pro Phe Pro Cys Ala Glu Cys Gly Lys Ala Phe Arg Glu
                470                 475                 480
Ser Ser Gln Leu Leu Gln His Gln Arg Thr His Thr Gly Glu Arg
                485                 490                 495
Pro Phe Glu Cys Ala Glu Cys Gly Gln Ala Phe Val Met Gly Ser
                500                 505                 510
Tyr Leu Ala Glu His Arg Arg Val His Thr Gly Glu Lys Pro His
                515                 520                 525
Ala Cys Ala Gln Cys Gly Lys Ala Phe Ser Gln Arg Ser Asn Leu
                530                 535                 540
Leu Ser His Arg Arg Thr His Ser Gly Ala Lys Pro Phe Ala Cys
                545                 550                 555
Ala Asp Cys Gly Lys Ala Phe Arg Gly Ser Ser Gly Leu Ala His
                560                 565                 570
His Arg Leu Ser His Thr Gly Glu Arg Pro Phe Ala Cys Ala Glu
                575                 580                 585
Cys Gly Lys Ala Phe Arg Gly Ser Ser Glu Leu Arg Gln His Gln
                590                 595                 600
Arg Leu His Ser Gly Glu Arg Pro Phe Val Cys Ala His Cys Ser
                605                 610                 615
Lys Ala Phe Val Arg Lys Ser Glu Leu Leu Ser His Arg Arg Thr
                620                 625                 630
His Thr Gly Glu Arg Pro Tyr Ala Cys Gly Glu Cys Gly Lys Pro
                635                 640                 645
Phe Ser His Arg Cys Asn Leu Asn Glu His Gln Lys Arg His Gly
                650                 655                 660
Gly Arg Ala Ala Pro
                665
```

<210> 22

<211> 452
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7488424CD1

<400> 22

Met	Asn	Ser	Gly	Ile	Ser	Gln	Val	Phe	Gln	Arg	Glu	Leu	Thr	Cys	
1				5					10						15
Pro	Ile	Cys	Leu	Asn	Tyr	Phe	Ile	Asp	Pro	Val	Thr	Ile	Asp	Cys	
				20					25						30
Gly	His	Ser	Phe	Cys	Arg	Pro	Cys	Phe	Tyr	Leu	Asn	Trp	Gln	Asp	
				35					40						45
Ile	Pro	Ile	Leu	Thr	Gln	Cys	Phe	Glu	Cys	Leu	Lys	Thr	Thr	Gln	
				50					55						60
Gln	Arg	Asn	Leu	Lys	Thr	Asn	Ile	Arg	Leu	Lys	Lys	Met	Ala	Ser	
				65					70						75
Arg	Ala	Arg	Lys	Ala	Ser	Leu	Trp	Leu	Phe	Leu	Ser	Ser	Glu	Glu	
				80					85						90
Gln	Met	Cys	Gly	Thr	His	Arg	Glu	Thr	Lys	Lys	Ile	Phe	Cys	Glu	
				95					100						105
Val	Asp	Arg	Ser	Leu	Leu	Cys	Leu	Leu	Cys	Ser	Ser	Ser	Leu	Glu	
				110					115						120
His	Arg	Tyr	His	Arg	His	Cys	Pro	Ala	Glu	Trp	Ala	Ala	Glu	Glu	
				125					130						135
His	Arg	Glu	Lys	Leu	Leu	Lys	Lys	Met	Gln	Ser	Leu	Trp	Glu	Lys	
				140					145						150
Val	Cys	Glu	Asn	Gln	Arg	Asn	Leu	Asn	Val	Glu	Thr	Thr	Arg	Ile	
				155					160						165
Ser	His	Trp	Lys	Asp	Tyr	Val	Asn	Val	Arg	Leu	Glu	Ala	Ile	Arg	
				170					175						180
Ala	Glu	Tyr	Gln	Lys	Met	Pro	Ala	Phe	His	His	Glu	Glu	Glu	Lys	
				185					190						195
His	Asn	Leu	Glu	Met	Leu	Lys	Lys	Lys	Gly	Lys	Glu	Ile	Phe	His	
				200					205						210
Arg	Leu	His	Leu	Ser	Lys	Ala	Lys	Met	Ala	His	Arg	Arg	Glu	Ile	
				215					220						225
Leu	Arg	Gly	Thr	Tyr	Ala	Glu	Leu	Met	Lys	Met	Cys	His	Lys	Pro	
				230					235						240
Asp	Val	Glu	Leu	Leu	Gln	Ala	Phe	Gly	Asp	Ile	Leu	His	Arg	Ser	
				245					250						255
Glu	Ser	Val	Leu	Leu	His	Met	Pro	Gln	Pro	Leu	Asn	Leu	Glu	Leu	
				260					265						270
Arg	Ala	Gly	Pro	Ile	Thr	Gly	Leu	Arg	Asp	Arg	Leu	Asn	Gln	Phe	
				275					280						285
Arg	Val	Asp	Ile	Thr	Leu	Pro	His	Asn	Glu	Ala	Asn	Ser	His	Ile	
				290					295						300
Phe	Arg	Arg	Gly	Asp	Leu	Arg	Ser	Ile	Cys	Ile	Gly	Cys	Asp	Arg	
				305					310						315
Gln	Asn	Ala	Pro	His	Ile	Thr	Ala	Thr	Pro	Thr	Ser	Phe	Leu	Ala	
				320					325						330
Trp	Gly	Ala	Gln	Thr	Phe	Thr	Ser	Gly	Lys	Tyr	Tyr	Trp	Glu	Val	
				335					340						345
His	Val	Gly	Asp	Ser	Trp	Asn	Trp	Ala	Phe	Gly	Val	Cys	Asn	Lys	
				350					355						360
Tyr	Trp	Lys	Gly	Thr	Asn	Gln	Asn	Gly	Asn	Ile	His	Gly	Glu	Glu	
				365					370						375
Gly	Leu	Phe	Ser	Leu	Gly	Cys	Val	Lys	Asn	Asp	Ile	Gln	Cys	Asn	
				380					385						390
Leu	Phe	Thr	Thr	Ser	Pro	Val	Thr	Leu	Gln	Tyr	Val	Pro	Arg	Pro	
				395					400						405

```
Thr Asn His Val Gly Leu Phe Leu Asp Cys Glu Ala Arg Thr Val
                410                 415                 420
Ser Phe Val Asp Val Asn Gln Ser Ser Pro Ile Tyr Thr Ile Pro
                425                 430                 435
Asn Cys Ser Phe Ser Pro Pro Leu Arg Pro Ile Phe Cys Cys Ile
                440                 445                 450
His Leu
```

<210> 23
<211> 387
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7487110CD1

<400> 23

```
Met Thr Met Glu Gly Ala Ser Gly Ser Ser Phe Gly Ile Asp Thr
  1                 5                  10                  15
Ile Leu Ser Ser Ala Ser Ser Gly Ser Pro Gly Met Met Asn Gly
                 20                  25                  30
Asp Phe Arg Pro Leu Gly Glu Ala Arg Thr Ala Asp Phe Arg Ser
                 35                  40                  45
Gln Ala Thr Pro Ser Pro Cys Ser Glu Ile Asp Thr Val Gly Thr
                 50                  55                  60
Ala Pro Ser Ser Pro Ile Ser Val Thr Met Glu Pro Pro Glu Pro
                 65                  70                  75
His Leu Val Ala Asp Ala Thr Gln His His His His Leu His His
                 80                  85                  90
Ser Gln Gln Pro Pro Pro Pro Ala Ala Ala Pro Thr Gln Ser Leu
                 95                 100                 105
Gln Pro Leu Pro Gln Gln Gln Gln Pro Leu Pro Pro Gln Gln Pro
                110                 115                 120
Pro Pro Pro Pro Pro Gln Gln Leu Gly Ser Ala Ala Ser Ala Pro
                125                 130                 135
Arg Thr Ser Thr Ser Ser Phe Leu Ile Lys Asp Ile Leu Gly Asp
                140                 145                 150
Ser Lys Pro Leu Ala Ala Cys Ala Pro Tyr Ser Thr Ser Val Ser
                155                 160                 165
Ser Pro His His Thr Pro Lys Gln Glu Ser Asn Ala Val His Glu
                170                 175                 180
Ser Phe Arg Pro Lys Leu Glu Gln Glu Asp Ser Lys Thr Lys Leu
                185                 190                 195
Asp Lys Arg Glu Asp Ser Gln Ser Asp Ile Lys Cys His Gly Thr
                200                 205                 210
Lys Glu Glu Gly Asp Arg Glu Ile Thr Ser Ser Arg Glu Ser Pro
                215                 220                 225
Pro Val Arg Ala Lys Lys Pro Arg Lys Ala Arg Thr Ala Phe Ser
                230                 235                 240
Asp His Gln Leu Asn Gln Leu Glu Arg Ser Phe Glu Arg Gln Lys
                245                 250                 255
Tyr Leu Ser Val Gln Asp Arg Met Asp Leu Ala Ala Ala Leu Asn
                260                 265                 270
Leu Thr Asp Thr Gln Val Lys Thr Trp Tyr Gln Asn Arg Arg Thr
                275                 280                 285
Lys Trp Lys Arg Gln Thr Ala Val Gly Leu Glu Leu Leu Ala Glu
                290                 295                 300
Ala Gly Asn Tyr Ser Ala Leu Gln Arg Met Phe Pro Ser Pro Tyr
                305                 310                 315
Phe Tyr His Pro Ser Leu Leu Gly Ser Met Asp Ser Thr Thr Ala
                320                 325                 330
```

```
Ala Ala Ala Ala Ala Ala Met Tyr Ser Ser Met Tyr Arg Thr Pro
                335                 340                 345
Pro Ala Pro His Pro Gln Leu Gln Arg Pro Leu Val Pro Arg Val
                350                 355                 360
Leu Ile His Gly Leu Gly Pro Gly Gly Gln Pro Ala Leu Asn Pro
                365                 370                 375
Leu Ser Ser Pro Ile Pro Gly Thr Pro His Pro Arg
                380                 385
```

<210> 24
<211> 255
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7495008CD1

<400> 24

```
Met Leu Arg Pro Gln Arg Pro Gly Asp Leu Gln Leu Gly Ala Ser
  1                 5                  10                  15
Leu Tyr Glu Leu Val Gly Tyr Arg Gln Pro Pro Ser Ser Ser Ser
                 20                  25                  30
Ser Ser Thr Ser Ser Thr Ser Ser Thr Ser Ser Ser Ser Thr Thr
                 35                  40                  45
Ala Pro Leu Leu Pro Lys Ala Ala Arg Glu Lys Pro Glu Ala Pro
                 50                  55                  60
Ala Glu Pro Pro Gly Pro Gly Pro Gly Ser Gly Ala His Pro Gly
                 65                  70                  75
Gly Ser Ala Arg Pro Asp Ala Lys Glu Glu Gln Gln Gln Gln Leu
                 80                  85                  90
Arg Arg Lys Ile Asn Ser Arg Glu Arg Lys Arg Met Gln Asp Leu
                 95                 100                 105
Asn Leu Ala Met Asp Ala Leu Arg Glu Val Ile Leu Pro Tyr Ser
                110                 115                 120
Ala Ala His Cys Gln Gly Ala Pro Gly Arg Lys Leu Ser Lys Ile
                125                 130                 135
Ala Thr Leu Leu Leu Ala Arg Asn Tyr Ile Leu Leu Leu Gly Ser
                140                 145                 150
Ser Leu Gln Glu Leu Arg Arg Ala Leu Gly Glu Gly Ala Gly Pro
                155                 160                 165
Ala Ala Pro Arg Leu Leu Leu Ala Gly Leu Pro Leu Leu Ala Ala
                170                 175                 180
Ala Pro Gly Ser Val Leu Leu Ala Pro Gly Ala Val Gly Pro Pro
                185                 190                 195
Asp Ala Leu Arg Pro Ala Lys Tyr Leu Ser Leu Ala Leu Asp Glu
                200                 205                 210
Pro Pro Cys Gly Gln Phe Ala Leu Pro Gly Gly Gly Ala Gly Gly
                215                 220                 225
Pro Gly Leu Cys Thr Cys Ala Val Cys Lys Phe Pro His Leu Val
                230                 235                 240
Pro Ala Ser Leu Gly Leu Ala Ala Val Gln Ala Gln Phe Ser Lys
                245                 250                 255
```

<210> 25
<211> 334
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7073515CD1

<400> 25
Met Phe Gly Lys Pro Asp Lys Met Asp Val Arg Cys His Ser Asp
1 5 10 15
Ala Glu Ala Ala Arg Val Ser Lys Asn Ala His Lys Glu Ser Arg
20 25 30
Glu Ser Lys Gly Ala Glu Gly Asn Leu Pro Ala Ala Phe Leu Lys
35 40 45
Glu Pro Gln Gly Ala Phe Ser Ala Ser Gly Ala Ala Glu Asp Cys
50 55 60
Asn Lys Ser Lys Ser Asn Ser Ala Ala Asp Pro Asp Tyr Cys Arg
65 70 75
Arg Ile Leu Val Arg Asp Ala Lys Gly Ser Ile Arg Glu Ile Ile
80 85 90
Leu Pro Lys Gly Leu Asp Leu Asp Arg Pro Lys Arg Thr Arg Thr
95 100 105
Ser Phe Thr Ala Glu Gln Leu Tyr Arg Leu Glu Met Glu Phe Gln
110 115 120
Arg Cys Gln Tyr Val Val Gly Arg Glu Arg Thr Glu Leu Ala Arg
125 130 135
Gln Leu Asn Leu Ser Glu Thr Gln Val Lys Val Trp Phe Gln Asn
140 145 150
Arg Arg Thr Lys Gln Lys Lys Asp Gln Gly Lys Asp Ser Glu Leu
155 160 165
Arg Ser Val Val Ser Glu Thr Ala Ala Thr Cys Ser Val Leu Arg
170 175 180
Leu Leu Glu Gln Gly Arg Leu Leu Ser Pro Pro Gly Leu Pro Ala
185 190 195
Leu Leu Pro Pro Cys Ala Thr Gly Ala Leu Gly Ser Ala Leu Arg
200 205 210
Gly Pro Ser Leu Pro Ala Leu Gly Ala Gly Ala Ala Ala Gly Ser
215 220 225
Ala Ala Ala Ala Ala Ala Ala Ala Pro Gly Pro Ala Gly Ala Ala
230 235 240
Ser Pro His Pro Pro Ala Val Gly Gly Ala Pro Gly Pro Gly Pro
245 250 255
Ala Gly Pro Gly Gly Leu His Ala Cys Ala Pro Ala Ala Gly His
260 265 270
Ser Leu Phe Ser Leu Pro Val Pro Ser Leu Leu Gly Ser Val Ala
275 280 285
Ser Arg Leu Ser Ser Ala Pro Leu Thr Met Ala Gly Ser Leu Ala
290 295 300
Gly Asn Leu Gln Glu Leu Ser Ala Arg Tyr Leu Ser Ser Ser Ala
305 310 315
Phe Glu Pro Tyr Ser Arg Thr Asn Asn Lys Glu Gly Ala Glu Lys
320 325 330
Lys Ala Leu Asp

<210> 26
<211> 262
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3356640CD1

<400> 26
Met Lys Arg His Glu Met Val Val Ala Lys His Ser Ala Leu Cys
1 5 10 15
Ser Arg Phe Ala Gln Asp Leu Trp Leu Glu Gln Asn Ile Lys Asp
20 25 30
Ser Phe Gln Lys Val Thr Leu Ser Arg Tyr Gly Lys Tyr Gly His

```
                  35                  40                  45
Lys Asn Leu Gln Leu Arg Lys Gly Cys Lys Ser Val Asp Glu Cys
                  50                  55                  60
Lys Glu His Gln Gly Gly Tyr Asn Gly Leu Asn Gln Cys Leu Lys
                  65                  70                  75
Ile Thr Thr Ser Lys Ile Phe Gln Cys Asn Lys Tyr Val Lys Val
                  80                  85                  90
Met His Lys Phe Ser Asn Ser Asn Arg His Lys Ile Arg His Thr
                  95                 100                 105
Glu Asn Lys His Phe Arg Cys Lys Glu Cys Asp Lys Ser Leu Cys
                 110                 115                 120
Met Leu Ser Arg Leu Thr Gln His Lys Lys Ile His Thr Arg Glu
                 125                 130                 135
Asn Phe Tyr Lys Cys Glu Glu Cys Gly Lys Thr Phe Asn Trp Ser
                 140                 145                 150
Thr Asn Leu Ser Lys Pro Lys Lys Ile His Thr Gly Glu Lys Pro
                 155                 160                 165
Tyr Lys Cys Glu Val Cys Gly Lys Ala Phe His Gln Ser Ser Ile
                 170                 175                 180
Leu Thr Lys His Lys Ile Ile Arg Thr Gly Glu Lys Pro Tyr Lys
                 185                 190                 195
Cys Ala His Cys Gly Lys Ala Phe Lys Gln Ser Ser His Leu Thr
                 200                 205                 210
Arg His Lys Ile Ile His Thr Glu Glu Lys Pro Tyr Lys Cys Glu
                 215                 220                 225
Gln Cys Gly Lys Val Phe Lys Gln Ser Pro Thr Leu Thr Lys His
                 230                 235                 240
Gln Ile Ile Tyr Thr Gly Glu Glu Pro Tyr Lys Cys Glu Glu Cys
                 245                 250                 255
Gly Lys Ala Phe Asn Leu Ser
                 260

<210> 27
<211> 509
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2015706CD1

<400> 27
Met Ala Leu Ser Gln Gly Leu Leu Thr Phe Arg Asp Val Ala Ile
  1                 5                  10                  15
Glu Phe Ser Gln Glu Glu Trp Lys Cys Leu Asp Pro Ala Gln Arg
                  20                  25                  30
Thr Leu Tyr Arg Asp Val Met Leu Glu Asn Tyr Arg Asn Leu Val
                  35                  40                  45
Ser Leu Asp Ile Ser Ser Arg Cys Met Met Asn Thr Leu Ser Ser
                  50                  55                  60
Thr Gly Gln Gly Asn Thr Glu Val Ile His Thr Gly Thr Leu Gln
                  65                  70                  75
Arg Gln Ala Ser Tyr His Ile Gly Ala Phe Cys Ser Gln Glu Ile
                  80                  85                  90
Glu Lys Asp Ile His Asp Phe Val Phe Gln Trp Gln Glu Asp Glu
                  95                 100                 105
Thr Asn Asp His Glu Ala Pro Met Thr Glu Ile Lys Lys Leu Thr
                 110                 115                 120
Ser Ser Thr Asp Arg Tyr Asp Gln Arg His Ala Gly Asn Lys Pro
                 125                 130                 135
Ile Lys Gly Gln Leu Glu Ser Arg Phe His Leu His Leu Arg Arg
                 140                 145                 150
His Arg Arg Ile His Thr Gly Glu Lys Pro Tyr Lys Cys Glu Glu
```

```
                155                 160                 165
Cys Glu Lys Val Phe Ser Cys Lys Ser His Leu Glu Ile His Arg
                170                 175                 180
Ile Ile His Thr Gly Glu Lys Pro Tyr Lys Cys Lys Val Cys Asp
                185                 190                 195
Lys Ala Phe Lys His Asp Ser His Leu Ala Lys His Thr Arg Ile
                200                 205                 210
His Arg Gly Asp Lys His Tyr Thr Cys Asn Glu Cys Gly Lys Val
                215                 220                 225
Phe Asp Gln Lys Ala Thr Leu Ala Cys His His Arg Ser His Thr
                230                 235                 240
Gly Glu Lys Pro Tyr Lys Cys Asn Glu Cys Gly Lys Thr Phe Ser
                245                 250                 255
Gln Thr Ser His Leu Val Tyr His His Arg Leu His Thr Gly Glu
                260                 265                 270
Lys Pro Tyr Lys Cys Asn Glu Cys Gly Lys Thr Phe Ala Arg Asn
                275                 280                 285
Ser Val Leu Val Ile His Lys Ala Val His Thr Ala Glu Lys Pro
                290                 295                 300
Tyr Lys Cys Asn Glu Cys Gly Lys Val Phe Lys Gln Arg Ala Thr
                305                 310                 315
Leu Ala Gly His Arg Arg Val His Thr Gly Glu Lys Pro Tyr Arg
                320                 325                 330
Cys Glu Glu Cys Asp Lys Val Phe Ser Arg Lys Ser His Leu Glu
                335                 340                 345
Arg His Arg Arg Ile His Thr Gly Glu Lys Pro Tyr Lys Cys Lys
                350                 355                 360
Val Cys Asp Lys Ala Phe Arg Ser Asp Ser Arg Leu Ala Glu His
                365                 370                 375
Gln Arg Val His Thr Gly Glu Arg Pro Tyr Thr Cys Asn Glu Cys
                380                 385                 390
Gly Lys Val Phe Ser Thr Lys Ala Tyr Leu Ala Cys His Gln Lys
                395                 400                 405
Leu His Thr Gly Glu Lys Leu Tyr Glu Cys Glu Glu Cys Asp Lys
                410                 415                 420
Val Tyr Ile Arg Lys Ser His Leu Glu Arg His Arg Arg Ile His
                425                 430                 435
Thr Gly Glu Lys Pro His Lys Cys Gly Asp Cys Gly Lys Ala Phe
                440                 445                 450
Asn Ser Pro Ser His Leu Ile Arg His Gln Arg Ile His Thr Gly
                455                 460                 465
Gln Lys Ser Tyr Lys Cys His Gln Cys Gly Lys Val Phe Ser Leu
                470                 475                 480
Arg Ser Leu Leu Ala Glu His Gln Lys Ile Pro Phe Gly Asp Asn
                485                 490                 495
Cys Phe Lys Cys Asn Glu Tyr Ser Lys Pro Ser Ser Ile Asn
                500                 505
```

<210> 28
<211> 310
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6920755CD1

<400> 28

```
Met Ser Gln Gln Leu Lys Lys Arg Ala Lys Thr Arg His Gln Lys
  1               5                  10                  15
Gly Leu Gly Gly Arg Ala Pro Ser Gly Ala Lys Pro Arg Gln Gly
                20                  25                  30
Lys Ser Ser Gln Asp Leu Gln Ala Glu Ile Glu Pro Val Ser Ala
```

```
                  35                  40                  45
Val Trp Ala Leu Cys Asp Gly Tyr Val Cys Tyr Glu Pro Gly Pro
                  50                  55                  60
Gln Ala Leu Gly Gly Asp Asp Phe Ser Asp Cys Tyr Ile Glu Cys
                  65                  70                  75
Val Ile Arg Gly Glu Phe Ser Gln Pro Ile Leu Glu Glu Asp Ser
                  80                  85                  90
Leu Phe Glu Ser Leu Glu Tyr Leu Lys Lys Gly Ser Glu Gln Gln
                  95                 100                 105
Leu Ser Gln Lys Val Phe Glu Ala Ser Ser Leu Glu Cys Ser Leu
                 110                 115                 120
Glu Tyr Met Lys Lys Gly Val Lys Lys Glu Leu Pro Gln Lys Ile
                 125                 130                 135
Val Gly Glu Asn Ser Leu Glu Tyr Ser Glu Tyr Met Thr Gly Lys
                 140                 145                 150
Lys Leu Pro Pro Gly Gly Ile Pro Gly Ile Asp Leu Ser Asp Pro
                 155                 160                 165
Lys Gln Leu Ala Glu Phe Ala Arg Lys Lys Pro Pro Ile Asn Lys
                 170                 175                 180
Glu Tyr Asp Ser Leu Ser Ala Ile Ala Cys Pro Gln Ser Gly Cys
                 185                 190                 195
Thr Arg Lys Leu Arg Asp Arg Ala Ala Leu Arg Lys His Leu Leu
                 200                 205                 210
Ile His Gly Pro Arg Asp His Val Cys Ala Glu Cys Gly Lys Ala
                 215                 220                 225
Phe Val Glu Ser Ser Lys Leu Lys Arg His Phe Leu Val His Thr
                 230                 235                 240
Gly Glu Lys Pro Phe Arg Cys Thr Phe Glu Gly Cys Gly Lys Arg
                 245                 250                 255
Phe Ser Leu Asp Phe Asn Leu Arg Thr His Val Arg Ile His Thr
                 260                 265                 270
Gly Glu Lys Arg Phe Val Cys Pro Phe Gln Gly Cys Asn Arg Arg
                 275                 280                 285
Phe Ile Gln Ser Asn Asn Leu Lys Ala His Ile Leu Thr His Ala
                 290                 295                 300
Asn Thr Asn Lys Asn Glu Gln Glu Gly Lys
                 305                 310
```

<210> 29
<211> 402
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 444179CD1

```
<400> 29
Met Ala Ala Val Ile Leu Pro Ser Thr Ala Ala Pro Ser Ser Leu
  1               5                  10                  15
Phe Pro Ala Ser Gln Gln Lys Gly His Thr Gln Gly Gly Glu Leu
                 20                  25                  30
Val Asn Glu Leu Leu Thr Ser Trp Leu Arg Gly Leu Val Thr Phe
                 35                  40                  45
Glu Asp Val Ala Val Glu Phe Thr Gln Glu Glu Trp Ala Leu Leu
                 50                  55                  60
Asp Pro Ala Gln Arg Thr Leu Tyr Arg Asp Val Met Leu Glu Asn
                 65                  70                  75
Cys Arg Asn Leu Ala Ser Leu Gly Cys Arg Val Asn Lys Pro Ser
                 80                  85                  90
Leu Ile Ser Gln Leu Glu Gln Asp Lys Lys Val Val Thr Glu Glu
                 95                 100                 105
Arg Gly Ile Leu Pro Ser Thr Cys Pro Asp Leu Glu Thr Leu Leu
```

```
                  110                 115                 120
Lys Ala Lys Trp Leu Thr Pro Lys Lys Asn Val Phe Arg Lys Glu
                  125                 130                 135
Gln Ser Lys Gly Val Lys Thr Glu Arg Ser His Arg Gly Val Lys
                  140                 145                 150
Leu Asn Glu Cys Asn Gln Cys Phe Lys Val Phe Ser Thr Lys Ser
                  155                 160                 165
Asn Leu Thr Gln His Lys Arg Ile His Thr Gly Glu Lys Pro Tyr
                  170                 175                 180
Asp Cys Ser Gln Cys Gly Lys Ser Phe Ser Ser Arg Ser Tyr Leu
                  185                 190                 195
Thr Ile His Lys Arg Ile His Asn Gly Glu Lys Pro Tyr Glu Cys
                  200                 205                 210
Asn His Cys Gly Lys Ala Phe Ser Asp Pro Ser Ser Leu Arg Leu
                  215                 220                 225
His Leu Arg Ile His Thr Gly Glu Lys Pro Tyr Glu Cys Asn Gln
                  230                 235                 240
Cys Phe His Val Phe Arg Thr Ser Cys Asn Leu Lys Ser His Lys
                  245                 250                 255
Arg Ile His Thr Gly Glu Asn His His Glu Cys Asn Gln Cys Gly
                  260                 265                 270
Lys Ala Phe Ser Thr Arg Ser Ser Leu Thr Gly His Asn Ser Ile
                  275                 280                 285
His Thr Gly Glu Lys Pro Tyr Glu Cys His Asp Cys Gly Lys Thr
                  290                 295                 300
Phe Arg Lys Ser Ser Tyr Leu Thr Gln His Val Arg Thr His Thr
                  305                 310                 315
Gly Glu Lys Pro Tyr Glu Cys Asn Glu Cys Gly Lys Ser Phe Ser
                  320                 325                 330
Ser Ser Phe Ser Leu Thr Val His Lys Arg Ile His Thr Gly Glu
                  335                 340                 345
Lys Pro Tyr Glu Cys Ser Asp Cys Gly Lys Ala Phe Asn Asn Leu
                  350                 355                 360
Ser Ala Val Lys Lys His Leu Arg Thr His Thr Gly Glu Lys Pro
                  365                 370                 375
Tyr Glu Cys Asn His Cys Gly Lys Ser Phe Thr Ser Asn Ser Tyr
                  380                 385                 390
Leu Ser Val His Lys Arg Ile His Asn Arg Trp Ile
                  395                 400
```

<210> 30
<211> 602
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5628380CD1

<400> 30

```
Met Ser Asn Glu Leu Asp Phe Arg Ser Val Arg Leu Leu Lys Asn
  1                 5                  10                  15
Asp Pro Val Asn Leu Gln Lys Phe Ser Tyr Thr Ser Glu Asp Glu
                  20                  25                  30
Ala Trp Lys Thr Tyr Leu Glu Asn Pro Leu Thr Ala Ala Thr Lys
                  35                  40                  45
Ala Met Met Arg Val Asn Gly Asp Asp Asp Ser Val Ala Ala Leu
                  50                  55                  60
Ser Phe Leu Tyr Asp Tyr Tyr Met Gly Pro Lys Glu Lys Arg Ile
                  65                  70                  75
Leu Ser Ser Ser Thr Gly Gly Arg Asn Asp Gln Gly Lys Arg Tyr
                  80                  85                  90
Tyr His Gly Met Glu Tyr Glu Thr Asp Leu Thr Pro Leu Glu Ser
```

```
                    95                  100                 105
Pro Thr His Leu Met Lys Phe Leu Thr Glu Asn Val Ser Gly Thr
                   110                 115                 120
Pro Glu Tyr Pro Asp Leu Leu Lys Lys Asn Asn Leu Met Ser Leu
                   125                 130                 135
Glu Gly Ala Leu Pro Thr Pro Gly Lys Ala Ala Pro Leu Pro Ala
                   140                 145                 150
Gly Pro Ser Lys Leu Glu Ala Gly Ser Val Asp Ser Tyr Leu Leu
                   155                 160                 165
Pro Thr Thr Asp Met Tyr Asp Asn Gly Ser Leu Asn Ser Leu Phe
                   170                 175                 180
Glu Ser Ile His Gly Val Pro Pro Thr Gln Arg Trp Gln Pro Asp
                   185                 190                 195
Ser Thr Phe Lys Asp Asp Pro Gln Glu Ser Met Leu Phe Pro Asp
                   200                 205                 210
Ile Leu Lys Thr Ser Pro Glu Pro Pro Cys Pro Glu Asp Tyr Pro
                   215                 220                 225
Ser Leu Lys Ser Asp Phe Glu Tyr Thr Leu Gly Ser Pro Lys Ala
                   230                 235                 240
Ile His Ile Lys Ser Gly Glu Ser Pro Met Ala Tyr Leu Asn Lys
                   245                 250                 255
Gly Gln Phe Tyr Pro Val Thr Leu Arg Thr Pro Ala Gly Gly Lys
                   260                 265                 270
Gly Leu Ala Leu Ser Ser Asn Lys Val Lys Ser Val Val Met Val
                   275                 280                 285
Val Phe Asp Asn Glu Lys Val Pro Val Glu Gln Leu Arg Phe Trp
                   290                 295                 300
Lys His Trp His Ser Arg Gln Pro Thr Ala Lys Gln Arg Val Ile
                   305                 310                 315
Asp Val Ala Asp Cys Lys Glu Asn Phe Asn Thr Val Glu His Ile
                   320                 325                 330
Glu Glu Val Ala Tyr Asn Ala Leu Ser Phe Val Trp Asn Val Asn
                   335                 340                 345
Glu Glu Ala Lys Val Phe Ile Gly Val Asn Cys Leu Ser Thr Asp
                   350                 355                 360
Phe Ser Ser Gln Lys Gly Val Lys Gly Val Pro Leu Asn Leu Gln
                   365                 370                 375
Ile Asp Thr Tyr Asp Cys Gly Leu Gly Thr Glu Arg Leu Val His
                   380                 385                 390
Arg Ala Val Cys Gln Ile Lys Ile Phe Cys Asp Lys Gly Ala Glu
                   395                 400                 405
Arg Lys Met Arg Asp Asp Glu Arg Lys Gln Phe Arg Arg Lys Val
                   410                 415                 420
Lys Cys Pro Asp Ser Ser Asn Ser Gly Val Lys Gly Cys Leu Leu
                   425                 430                 435
Ser Gly Phe Arg Gly Asn Glu Thr Thr Tyr Leu Arg Pro Glu Thr
                   440                 445                 450
Asp Leu Glu Thr Pro Pro Val Leu Phe Ile Pro Asn Val His Phe
                   455                 460                 465
Ser Ser Leu Gln Arg Ser Gly Gly Ala Ala Pro Ser Ala Gly Pro
                   470                 475                 480
Ser Ser Ser Asn Arg Leu Pro Leu Lys Arg Thr Cys Ser Pro Phe
                   485                 490                 495
Thr Glu Glu Phe Glu Pro Leu Pro Ser Lys Gln Ala Lys Glu Gly
                   500                 505                 510
Asp Leu Gln Arg Val Leu Leu Tyr Val Arg Arg Glu Thr Glu Glu
                   515                 520                 525
Val Phe Asp Ala Leu Met Leu Lys Thr Pro Asp Leu Lys Gly Leu
                   530                 535                 540
Arg Asn Ala Ile Ser Glu Lys Tyr Gly Phe Pro Glu Glu Asn Ile
                   545                 550                 555
Tyr Lys Val Tyr Lys Lys Cys Lys Arg Gly Ile Leu Val Asn Met
                   560                 565                 570
```

```
Asp Asn Asn Ile Ile Gln His Tyr Ser Asn His Val Ala Phe Leu
                575                 580                 585
Leu Asp Met Gly Glu Leu Asp Gly Lys Ile Gln Ile Ile Leu Lys
                590                 595                 600
Glu Leu
```

<210> 31
<211> 816
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7493789CD1

<400> 31

```
Met Glu Ile Gly Ser Ala Gly Pro Ala Gly Ala Gln Pro Leu Leu
  1               5                  10                  15
Met Val Pro Arg Arg Pro Gly Tyr Gly Thr Met Gly Lys Pro Ile
                20                  25                  30
Lys Leu Leu Ala Asn Cys Phe Gln Val Glu Ile Pro Lys Ile Asp
                35                  40                  45
Val Tyr Leu Tyr Glu Val Asp Ile Lys Pro Asp Lys Cys Pro Arg
                50                  55                  60
Arg Val Asn Arg Glu Val Val Asp Ser Met Val Gln His Phe Lys
                65                  70                  75
Val Thr Ile Phe Gly Asp Arg Arg Pro Val Tyr Asp Gly Lys Arg
                80                  85                  90
Ser Leu Tyr Thr Ala Asn Pro Leu Pro Val Ala Thr Thr Gly Val
                95                 100                 105
Asp Leu Asp Val Thr Leu Pro Gly Glu Gly Gly Lys Asp Arg Pro
               110                 115                 120
Phe Lys Val Ser Ile Lys Phe Val Ser Arg Tyr Thr Pro Val Gly
               125                 130                 135
Arg Ser Phe Phe Ser Ala Pro Glu Gly Tyr Asp His Pro Leu Gly
               140                 145                 150
Gly Gly Arg Glu Val Trp Phe Gly Phe His Gln Ser Val Arg Pro
               155                 160                 165
Ala Met Trp Lys Met Met Leu Asn Ile Asp Val Ser Ala Thr Ala
               170                 175                 180
Phe Tyr Lys Ala Gln Pro Val Ile Gln Phe Met Cys Glu Val Leu
               185                 190                 195
Asp Ile His Asn Ile Asp Glu Gln Pro Arg Pro Leu Thr Asp Ser
               200                 205                 210
His Arg Val Lys Phe Thr Lys Glu Ile Lys Gly Leu Lys Val Glu
               215                 220                 225
Val Thr His Cys Gly Thr Met Arg Arg Lys Tyr Arg Val Cys Asn
               230                 235                 240
Val Thr Arg Arg Pro Ala Ser His Gln Thr Phe Pro Leu Gln Leu
               245                 250                 255
Glu Asn Gly Gln Thr Val Glu Arg Thr Val Ala Gln Tyr Phe Arg
               260                 265                 270
Glu Lys Tyr Thr Leu Gln Leu Lys Tyr Pro His Leu Pro Cys Leu
               275                 280                 285
Gln Val Gly Gln Glu Gln Lys His Thr Tyr Leu Pro Leu Glu Val
               290                 295                 300
Cys Asn Ile Val Ala Gly Gln Arg Cys Ile Lys Lys Leu Thr Asp
               305                 310                 315
Asn Gln Thr Ser Thr Met Ile Lys Ala Thr Ala Arg Ser Ala Pro
               320                 325                 330
Asp Arg Gln Glu Glu Ile Ser Arg Leu Val Arg Ser Ala Asn Tyr
               335                 340                 345
```

```
Glu Thr Asp Pro Phe Val Gln Glu Phe Gln Phe Lys Val Arg Asp
                350                 355                 360
Glu Met Ala His Val Thr Gly Arg Val Leu Pro Ala Pro Met Leu
                365                 370                 375
Gln Tyr Gly Gly Arg Asn Arg Thr Val Ala Thr Pro Ser His Gly
                380                 385                 390
Val Trp Asp Met Arg Gly Lys Gln Phe His Thr Gly Val Glu Ile
                395                 400                 405
Lys Met Trp Ala Ile Ala Cys Phe Ala Thr Gln Arg Gln Cys Arg
                410                 415                 420
Glu Glu Ile Leu Lys Gly Phe Thr Asp Gln Leu Arg Lys Ile Ser
                425                 430                 435
Lys Asp Ala Gly Met Pro Ile Gln Gly Gln Pro Cys Phe Cys Lys
                440                 445                 450
Tyr Ala Gln Gly Ala Asp Ser Val Glu Pro Met Phe Arg His Leu
                455                 460                 465
Lys Asn Thr Tyr Ser Gly Leu Gln Leu Ile Ile Val Ile Leu Pro
                470                 475                 480
Gly Lys Thr Pro Val Tyr Ala Glu Val Lys Arg Val Gly Asp Thr
                485                 490                 495
Leu Leu Gly Met Ala Thr Gln Cys Val Gln Val Lys Asn Val Ile
                500                 505                 510
Lys Thr Ser Pro Gln Thr Leu Ser Asn Leu Cys Leu Lys Ile Asn
                515                 520                 525
Val Lys Leu Gly Gly Ile Asn Asn Ile Leu Val Pro His Gln Arg
                530                 535                 540
Pro Ser Val Phe Gln Gln Pro Val Ile Phe Leu Gly Ala Asp Val
                545                 550                 555
Thr His Pro Pro Ala Gly Asp Gly Lys Lys Pro Ser Ile Ala Ala
                560                 565                 570
Val Val Gly Ser Met Asp Ala His Pro Ser Arg Tyr Cys Ala Thr
                575                 580                 585
Val Arg Val Gln Arg Pro Arg Gln Glu Ile Ile Gln Asp Leu Ala
                590                 595                 600
Ser Met Val Arg Glu Leu Leu Ile Gln Phe Tyr Lys Ser Thr Arg
                605                 610                 615
Phe Lys Pro Thr Arg Ile Ile Phe Tyr Arg Asp Gly Val Ser Glu
                620                 625                 630
Gly Gln Phe Arg Gln Val Leu Tyr Tyr Glu Leu Leu Ala Ile Arg
                635                 640                 645
Glu Ala Cys Ile Ser Leu Glu Lys Asp Tyr Gln Pro Gly Ile Thr
                650                 655                 660
Tyr Ile Val Val Gln Lys Arg His His Thr Arg Leu Phe Cys Ala
                665                 670                 675
Asp Arg Thr Glu Arg Val Gly Arg Ser Gly Asn Ile Pro Ala Gly
                680                 685                 690
Thr Thr Val Asp Thr Asp Ile Thr His Pro Tyr Glu Phe Asp Phe
                695                 700                 705
Tyr Leu Cys Ser His Ala Gly Ile Gln Gly Thr Ser Arg Pro Ser
                710                 715                 720
His Tyr His Val Leu Trp Asp Asp Asn Cys Phe Thr Ala Asp Glu
                725                 730                 735
Leu Gln Leu Leu Thr Tyr Gln Leu Cys His Thr Tyr Val Arg Cys
                740                 745                 750
Thr Arg Ser Val Ser Ile Pro Ala Pro Ala Tyr Tyr Ala His Leu
                755                 760                 765
Val Ala Phe Arg Ala Arg Tyr His Leu Val Asp Lys Glu His Asp
                770                 775                 780
Ser Ala Glu Gly Ser His Val Ser Gly Gln Ser Asn Gly Arg Asp
                785                 790                 795
Pro Gln Ala Leu Ala Lys Ala Val Gln Ile His Gln Asp Thr Leu
                800                 805                 810
Arg Thr Met Tyr Phe Ala
```

815

<210> 32
<211> 2248
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2075194CD1

<400> 32

```
Met Met Asp Ser Glu Asn Lys Pro Glu Asn Asp Glu Asp Glu Lys
  1               5                  10                  15
Ile Asn Lys Glu Ala Gln Asp Leu Thr Lys Leu Ser Ser His Asn
             20                  25                  30
Glu Asp Gly Gly Pro Val Ser Asp Val Ile Ala Ser Phe Pro Glu
             35                  40                  45
Asn Ser Met Gly Lys Arg Gly Phe Ser Glu Ser Ser Asn Ser Asp
             50                  55                  60
Ser Val Val Ile Gly Glu Asp Arg Asn Lys His Ala Ser Lys Arg
             65                  70                  75
Arg Lys Leu Asp Glu Ala Glu Pro Leu Lys Ser Gly Lys Gln Gly
             80                  85                  90
Ile Cys Arg Leu Glu Thr Ser Glu Ser Ser Val Thr Glu Gly Gly
             95                 100                 105
Ile Ala Leu Asp Glu Thr Gly Lys Glu Thr Phe Leu Ser Asp Cys
            110                 115                 120
Thr Val Gly Gly Thr Cys Leu Pro Asn Ala Leu Ser Pro Ser Cys
            125                 130                 135
Asn Phe Ser Thr Ile Asp Val Val Ser Leu Lys Thr Asp Thr Glu
            140                 145                 150
Lys Thr Ser Ala Gln Glu Met Val Ser Leu Asp Leu Glu Arg Glu
            155                 160                 165
Ser Pro Phe Pro Pro Lys Glu Ile Ser Val Ser Cys Thr Ile Gly
            170                 175                 180
Asn Val Asp Thr Val Leu Lys Cys Gln Ile Cys Gly His Leu Phe
            185                 190                 195
Ser Ser Cys Ser Asp Leu Glu Lys His Ala Glu Ser His Met Gln
            200                 205                 210
Gln Pro Lys Glu His Thr Cys Cys His Cys Ser His Lys Ala Glu
            215                 220                 225
Ser Ser Ser Ala Leu His Met His Ile Lys Gln Ala His Gly Pro
            230                 235                 240
Gln Lys Val Phe Ser Cys Asp Leu Cys Gly Phe Gln Cys Ser Glu
            245                 250                 255
Glu Asn Leu Leu Asn Ala His Tyr Leu Gly Lys Thr His Leu Arg
            260                 265                 270
Arg Gln Asn Leu Ala Ala Arg Gly Gly Phe Val Gln Ile Leu Thr
            275                 280                 285
Lys Gln Pro Phe Pro Lys Lys Pro Arg Thr Met Ala Thr Lys Asn
            290                 295                 300
Val His Ser Lys Pro Arg Thr Ser Lys Ser Ile Ala Lys Asn Ser
            305                 310                 315
Asp Ser Lys Gly Leu Arg Asn Val Gly Ser Thr Phe Lys Asp Phe
            320                 325                 330
Arg Gly Ser Ile Ser Lys Gln Ser Gly Ser Ser Ser Glu Leu Leu
            335                 340                 345
Val Glu Met Met Pro Ser Arg Asn Thr Leu Ser Gln Glu Val Glu
            350                 355                 360
Ile Val Glu Glu His Val Thr Ser Leu Gly Leu Ala Gln Asn Pro
            365                 370                 375
Glu Asn Gln Ser Arg Lys Leu Asp Thr Leu Val Thr Ser Glu Gly
```

```
                380                 385                 390
Leu Leu Glu Lys Leu Glu Ser Thr Lys Asn Thr Leu Gln Ala Ala
                395                 400                 405
His Gly Asn Ser Val Thr Ser Arg Pro Arg Pro Glu Arg Asn Ile
                410                 415                 420
Leu Val Leu Gly Asn Ser Phe Arg Arg Arg Ser Ser Thr Phe Thr
                425                 430                 435
Leu Lys Gly Gln Ala Lys Lys Arg Phe Asn Leu Leu Gly Ile Lys
                440                 445                 450
Arg Gly Thr Ser Glu Thr Gln Arg Met Tyr Met Lys His Leu Arg
                455                 460                 465
Thr Gln Met Lys Thr His Asp Ala Glu Ser Val Leu Lys His Leu
                470                 475                 480
Glu Ala Cys Ser Ser Val Gln Arg Val Cys Val Thr Thr Ser Glu
                485                 490                 495
Thr Gln Glu Ala Glu Gln Gly Gln Gly Ser Ala Arg Pro Pro Asp
                500                 505                 510
Ser Gly Leu His Ser Leu Thr Val Lys Pro Ala Ser Gly Ser Gln
                515                 520                 525
Thr Leu Cys Ala Cys Thr Asp Cys Gly Gln Val Ala Thr Asn Arg
                530                 535                 540
Thr Asp Leu Glu Ile His Val Lys Arg Cys His Ala Arg Glu Met
                545                 550                 555
Lys Phe Tyr Cys Arg Thr Cys Asp Phe Ser Ser Met Ser Arg Arg
                560                 565                 570
Asp Leu Asp Glu His Leu His Ser Asn Gln His Gln Gln Thr Ala
                575                 580                 585
Ser Val Leu Ser Cys Gln Cys Cys Ser Phe Ile Ser Leu Asp Glu
                590                 595                 600
Ile Asn Leu Arg Asp His Met Lys Glu Lys His Asn Met His Phe
                605                 610                 615
Leu Cys Thr Pro Cys Asn Leu Phe Phe Leu Ser Glu Lys Asp Val
                620                 625                 630
Glu Glu His Lys Ala Thr Glu Lys His Ile Asn Ser Leu Val Gln
                635                 640                 645
Pro Lys Thr Leu Gln Ser Ser Asn Ser Asp Leu Val Leu Gln Thr
                650                 655                 660
Leu Pro Leu Ser Thr Leu Glu Ser Glu Asn Ala Lys Glu Ser Met
                665                 670                 675
Asp Asp Ser Gly Lys Ala Ser Gln Glu Glu Pro Leu Lys Ser Arg
                680                 685                 690
Val Ser His Gly Asn Glu Val Arg His Ser Ser Lys Pro Gln Phe
                695                 700                 705
Gln Cys Lys Lys Cys Phe Tyr Lys Thr Arg Ser Ser Thr Val Leu
                710                 715                 720
Thr Arg His Ile Lys Leu Arg His Gly Gln Asp Tyr His Phe Leu
                725                 730                 735
Cys Lys Ala Cys Asn Leu Tyr Ser Leu Ser Lys Glu Gly Met Glu
                740                 745                 750
Lys His Ile Lys Arg Ser Lys His Leu Glu Asn Ala Lys Lys Asn
                755                 760                 765
Asn Ile Gly Leu Ser Phe Glu Glu Cys Ile Glu Arg Val Cys Ile
                770                 775                 780
Gly Ala Asn Asp Lys Lys Glu Glu Phe Asp Val Ser Gly Asn Gly
                785                 790                 795
Arg Ile Glu Gly His Ile Gly Val Gln Leu Gln Glu His Ser Tyr
                800                 805                 810
Leu Glu Lys Gly Met Leu Ala Ser Glu Glu Leu Ser Gln Ser Gly
                815                 820                 825
Gly Ser Thr Lys Asp Asp Glu Leu Ala Ser Thr Thr Thr Pro Lys
                830                 835                 840
Arg Gly Arg Pro Lys Gly Asn Ile Ser Arg Thr Cys Ser His Cys
                845                 850                 855
```

```
Gly Leu Leu Ala Ser Ser Ile Thr Asn Leu Thr Val His Ile Arg
                860                 865                 870
Arg Lys His Ser His Gln Tyr Ser Tyr Leu Cys Lys Val Cys Lys
                875                 880                 885
Tyr Tyr Thr Val Thr Lys Gly Asp Met Glu Arg His Cys Ala Thr
                890                 895                 900
Lys Lys His Lys Gly Arg Val Glu Ile Glu Ala Ser Gly Lys His
                905                 910                 915
Ser Ser Asp Ile Ile Val Gly Pro Glu Gly Gly Ser Leu Glu Ala
                920                 925                 930
Gly Lys Lys Asn Ala Gly Ser Ala Val Thr Met Ser Asp Glu His
                935                 940                 945
Ala Asn Lys Pro Ala Glu Ser Pro Thr Ser Val Leu Glu Lys Pro
                950                 955                 960
Asp Arg Gly Asn Ser Ile Glu Ala Glu Val Glu Asn Val Phe His
                965                 970                 975
Ser Leu Asp Gly Glu Val Asn Ser His Leu Leu Asp Lys Lys Glu
                980                 985                 990
Gln Ile Ser Ser Glu Pro Glu Asp Phe Ala Gln Pro Gly Asp Val
                995                1000                1005
Tyr Ser Gln Arg Asp Val Thr Gly Thr Gly Glu Asn Lys Cys Leu
               1010                1015                1020
His Cys Glu Phe Ser Ala His Ser Ser Ala Ser Leu Glu Leu His
               1025                1030                1035
Val Lys Arg Lys His Thr Lys Glu Phe Glu Phe Tyr Cys Met Ala
               1040                1045                1050
Cys Asp Tyr Tyr Ala Val Thr Arg Arg Glu Met Thr Arg His Ala
               1055                1060                1065
Ala Thr Glu Lys His Lys Met Lys Arg Gln Ser Tyr Leu Asn Ser
               1070                1075                1080
Ala Asn Val Glu Ala Gly Ser Ala Asp Met Ser Lys Asn Ile Ile
               1085                1090                1095
Met Pro Glu Glu Glu His Gln Gln Asn Ser Glu Glu Phe Gln Ile
               1100                1105                1110
Ile Ser Gly Gln Pro Ser Asp Thr Leu Lys Ser Arg Asn Ala Ala
               1115                1120                1125
Asp Cys Ser Ile Leu Asn Glu Asn Thr Asn Leu Asp Met Ser Lys
               1130                1135                1140
Val Leu Cys Ala Ala Asp Ser Val Glu Val Glu Thr Glu Glu Glu
               1145                1150                1155
Ser Asn Phe Asn Glu Asp His Ser Phe Cys Glu Thr Phe Gln Gln
               1160                1165                1170
Ala Pro Val Lys Asp Lys Val Arg Lys Pro Glu Glu Met Met Ser
               1175                1180                1185
Leu Thr Met Ser Ser Asn Tyr Gly Ser Pro Ser Arg Phe Gln Asn
               1190                1195                1200
Glu Asn Ser Gly Ser Ser Ala Leu Asn Cys Glu Thr Ala Lys Lys
               1205                1210                1215
Asn His Glu Ile Ser Asn Asp Ala Gly Glu Leu Arg Val His Cys
               1220                1225                1230
Glu Gly Glu Gly Gly Asn Ala Gly Asp Gly Gly Gly Val Val Pro
               1235                1240                1245
His Arg His Leu Cys Pro Val Thr Leu Asp Gly Glu Arg Ser Ala
               1250                1255                1260
Glu Ser Pro Val Leu Val Val Thr Arg Ile Thr Arg Glu Gln Gly
               1265                1270                1275
Asn Leu Glu Ser Gly Gly Gln Asn Arg Val Ala Arg Gly His Gly
               1280                1285                1290
Leu Glu Asp Leu Lys Gly Val Gln Glu Asp Pro Val Leu Gly Asn
               1295                1300                1305
Lys Glu Ile Leu Met Asn Ser Gln His Glu Thr Glu Phe Ile Leu
               1310                1315                1320
Glu Glu Asp Gly Pro Ala Ser Asp Ser Thr Val Glu Ser Ser Asp
```

```
                1325                1330                1335
Val Tyr Glu Thr Ile Ile Ser Ile Asp Asp Lys Gly Gln Ala Met
                1340                1345                1350
Tyr Ser Phe Gly Arg Phe Asp Ser Ser Ile Ile Arg Ile Lys Asn
                1355                1360                1365
Pro Glu Asp Gly Glu Leu Ile Asp Gln Ser Glu Glu Gly Leu Ile
                1370                1375                1380
Ala Thr Gly Val Arg Ile Ser Glu Leu Pro Leu Lys Asp Cys Ala
                1385                1390                1395
Gln Gly Val Lys Lys Lys Lys Ser Glu Gly Ser Ser Ile Gly Glu
                1400                1405                1410
Ser Thr Arg Ile Arg Cys Asp Asp Cys Gly Phe Leu Ala Asp Gly
                1415                1420                1425
Leu Ser Gly Leu Asn Val His Ile Ala Met Lys His Pro Thr Lys
                1430                1435                1440
Glu Lys His Phe His Cys Leu Leu Cys Gly Lys Ser Phe Tyr Thr
                1445                1450                1455
Glu Ser Asn Leu His Gln His Leu Ala Ser Ala Gly His Met Arg
                1460                1465                1470
Asn Glu Gln Ala Ser Val Glu Glu Leu Pro Glu Gly Gly Ala Thr
                1475                1480                1485
Phe Lys Cys Val Lys Cys Thr Glu Pro Phe Asp Ser Glu Gln Asn
                1490                1495                1500
Leu Phe Leu His Ile Lys Gly Gln His Glu Glu Leu Leu Arg Glu
                1505                1510                1515
Val Asn Lys Tyr Ile Val Glu Asp Thr Glu Gln Ile Asn Arg Glu
                1520                1525                1530
Arg Glu Glu Asn Gln Gly Asn Val Cys Lys Tyr Cys Gly Lys Met
                1535                1540                1545
Cys Arg Ser Ser Asn Ser Met Ala Phe Leu Ala His Ile Arg Thr
                1550                1555                1560
His Thr Gly Ser Lys Pro Phe Lys Cys Lys Ile Cys His Phe Ala
                1565                1570                1575
Thr Ala Gln Leu Gly Asp Ala Arg Asn His Val Lys Arg His Leu
                1580                1585                1590
Gly Met Arg Glu Tyr Lys Cys His Val Cys Gly Val Ala Phe Val
                1595                1600                1605
Met Lys Lys His Leu Asn Thr His Leu Leu Gly Lys His Gly Val
                1610                1615                1620
Gly Thr Pro Lys Glu Arg Lys Phe Thr Cys His Leu Cys Asp Arg
                1625                1630                1635
Ser Phe Thr Glu Lys Trp Ala Leu Asn Asn His Met Lys Leu His
                1640                1645                1650
Thr Gly Glu Lys Pro Phe Lys Cys Thr Trp Pro Thr Cys His Tyr
                1655                1660                1665
Ser Phe Leu Thr Ala Ser Ala Met Lys Asp His Tyr Arg Thr His
                1670                1675                1680
Thr Gly Glu Lys Ser Phe Leu Cys Asp Leu Cys Gly Phe Ala Gly
                1685                1690                1695
Gly Thr Arg His Ala Leu Thr Lys His Arg Arg Gln His Thr Gly
                1700                1705                1710
Glu Lys Pro Phe Lys Cys Asp Glu Cys Asn Phe Ala Ser Thr Thr
                1715                1720                1725
Gln Ser His Leu Thr Arg His Lys Arg Val His Thr Gly Glu Lys
                1730                1735                1740
Pro Tyr Arg Cys Pro Trp Cys Asp Tyr Arg Ser Asn Cys Ala Glu
                1745                1750                1755
Asn Ile Arg Lys His Ile Leu His Thr Gly Lys His Glu Gly Val
                1760                1765                1770
Lys Met Tyr Asn Cys Pro Lys Cys Asp Tyr Gly Thr Asn Val Pro
                1775                1780                1785
Val Glu Phe Arg Asn His Leu Lys Glu Gln His Pro Asp Ile Glu
                1790                1795                1800
```

```
Asn Pro Asp Leu Ala Tyr Leu His Ala Gly Ile Val Ser Lys Ser
                1805                1810                1815
Tyr Glu Cys Arg Leu Lys Gly Gln Gly Ala Thr Phe Val Glu Thr
                1820                1825                1830
Asp Ser Pro Phe Thr Ala Ala Ala Leu Ala Glu Glu Pro Leu Val
                1835                1840                1845
Lys Glu Lys Pro Leu Arg Ser Ser Arg Arg Pro Ala Pro Pro Pro
                1850                1855                1860
Glu Gln Val Gln Gln Val Ile Ile Phe Gln Gly Tyr Asp Gly Glu
                1865                1870                1875
Phe Ala Leu Asp Pro Ser Val Glu Glu Thr Ala Ala Ala Thr Leu
                1880                1885                1890
Gln Thr Leu Ala Met Ala Gly Gln Val Ala Arg Val Val His Ile
                1895                1900                1905
Thr Glu Asp Gly Gln Val Ile Ala Thr Ser Gln Ser Gly Ala His
                1910                1915                1920
Val Gly Ser Val Val Pro Gly Pro Ile Leu Pro Glu Gln Leu Ala
                1925                1930                1935
Asp Gly Ala Thr Gln Val Val Val Val Gly Gly Ser Met Glu Gly
                1940                1945                1950
His Gly Met Asp Glu Ser Leu Ser Pro Gly Gly Ala Val Ile Gln
                1955                1960                1965
Gln Val Thr Lys Gln Glu Ile Leu Asn Leu Ser Glu Ala Gly Val
                1970                1975                1980
Ala Pro Pro Glu Ala Ser Ser Ala Leu Asp Ala Leu Leu Cys Ala
                1985                1990                1995
Val Thr Glu Leu Gly Glu Val Glu Gly Arg Ala Gly Leu Glu Glu
                2000                2005                2010
Gln Gly Arg Pro Gly Ala Lys Asp Val Leu Ile Gln Leu Pro Gly
                2015                2020                2025
Gln Glu Val Ser His Val Ala Ala Asp Pro Glu Ala Pro Glu Ile
                2030                2035                2040
Gln Met Phe Pro Gln Ala Gln Glu Ser Pro Ala Ala Val Glu Val
                2045                2050                2055
Leu Thr Gln Val Val His Pro Ser Ala Ala Met Ala Ser Gln Glu
                2060                2065                2070
Arg Ala Gln Val Ala Phe Lys Lys Met Val Gln Gly Val Leu Gln
                2075                2080                2085
Phe Ala Val Cys Asp Thr Ala Ala Ala Gly Gln Leu Val Lys Asp
                2090                2095                2100
Gly Val Thr Gln Val Val Val Ser Glu Glu Gly Ala Val His Met
                2105                2110                2115
Val Ala Gly Glu Gly Ala Gln Ile Ile Met Gln Glu Ala Gln Gly
                2120                2125                2130
Glu His Met Asp Leu Val Glu Ser Asp Gly Glu Ile Ser Gln Ile
                2135                2140                2145
Ile Val Thr Glu Glu Leu Val Gln Ala Met Val Gln Glu Ser Ser
                2150                2155                2160
Gly Gly Phe Ser Glu Gly Thr Thr His Tyr Ile Leu Thr Glu Leu
                2165                2170                2175
Pro Pro Gly Val Gln Asp Glu Pro Gly Leu Tyr Ser His Thr Val
                2180                2185                2190
Leu Glu Thr Ala Asp Ser Gln Glu Leu Leu Gln Ala Gly Ala Thr
                2195                2200                2205
Leu Gly Thr Glu Ala Gly Ala Pro Ser Arg Ala Glu Gln Leu Ala
                2210                2215                2220
Ser Val Val Ile Tyr Thr Gln Glu Gly Ser Ser Ala Ala Ala Ala
                2225                2230                2235
Ile Gln Ser Gln Arg Glu Ser Ser Glu Leu Gln Glu Ala
                2240                2245
```

<210> 33
<211> 256

<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2801633CD1

<400> 33
Met Glu Leu Leu Thr Phe Lys Asp Val Ala Ile Glu Phe Ser Pro
1 5 10 15
Glu Glu Trp Lys Cys Leu Asp Ile Ser Gln Gln Asn Leu Tyr Arg
20 25 30
Asp Val Met Leu Glu Asn Tyr Arg Asn Leu Val Ser Leu Gly Val
35 40 45
Thr Ile Ser Asn Pro Asp Leu Val Thr Ser Leu Glu Gln Arg Lys
50 55 60
Glu Pro Tyr Asn Leu Lys Ile His Glu Thr Ala Ala Arg Pro Pro
65 70 75
Ala Val Cys Ser His Phe Thr Gln Asn Leu Trp Thr Val Gln Gly
80 85 90
Ile Glu Asp Ser Phe His Lys Leu Ile Pro Lys Gly His Glu Lys
95 100 105
Arg Gly His Glu Asn Leu Arg Lys Thr Cys Lys Ser Ile Asn Glu
110 115 120
Cys Lys Val Gln Lys Gly Gly Tyr Asn Arg Ile Asn Gln Cys Leu
125 130 135
Leu Thr Thr Gln Lys Lys Thr Ile Gln Ser Asn Ile Cys Val Lys
140 145 150
Val Phe His Lys Phe Ser Asn Ser Asn Lys Asp Lys Ile Arg Tyr
155 160 165
Thr Gly Asp Lys Thr Phe Lys Cys Lys Glu Cys Gly Lys Ser Phe
170 175 180
His Val Leu Ser Arg Leu Thr Gln His Lys Arg Ile His Thr Gly
185 190 195
Glu Asn Pro Tyr Thr Cys Glu Glu Cys Gly Lys Ala Phe Asn Trp
200 205 210
Ser Ser Ile Leu Thr Lys His Lys Arg Ile His Ala Arg Glu Lys
215 220 225
Phe Tyr Lys Cys Glu Glu Cys Gly Lys Gly Phe Thr Arg Ser Ser
230 235 240
His Leu Thr Lys His Lys Arg Ile His Thr Gly Glu Lys Leu Tyr
245 250 255
Thr

<210> 34
<211> 615
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7493525CD1

<400> 34
Met Asp Asp Leu Lys Tyr Gly Val Tyr Pro Leu Lys Glu Ala Ser
1 5 10 15
Gly Cys Pro Gly Ala Glu Arg Asn Leu Leu Val Tyr Ser Tyr Phe
20 25 30
Glu Lys Glu Thr Leu Thr Phe Arg Asp Val Ala Ile Glu Phe Ser
35 40 45
Leu Glu Glu Trp Glu Cys Leu Asn Pro Ala Gln Gln Asn Leu Tyr
50 55 60

```
Met Asn Val Met Leu Glu Asn Tyr Lys Asn Leu Val Phe Leu Ala
                65                  70                  75
Gly Val Ala Val Ser Lys Gln Asp Pro Val Thr Cys Leu Glu Gln
                80                  85                  90
Glu Lys Glu Pro Trp Asn Met Lys Arg His Glu Met Val Asp Glu
                95                 100                 105
Pro Pro Ala Met Cys Ser Tyr Phe Thr Lys Asp Leu Trp Pro Glu
               110                 115                 120
Gln Asp Ile Lys Asp Ser Phe Gln Gln Val Ile Leu Arg Arg Tyr
               125                 130                 135
Gly Lys Cys Glu His Glu Asn Leu Gln Leu Arg Lys Gly Ser Ala
               140                 145                 150
Ser Val Asp Glu Tyr Lys Val His Lys Glu Gly Tyr Asn Glu Leu
               155                 160                 165
Asn Gln Cys Leu Thr Thr Thr Gln Ser Lys Ile Phe Pro Cys Asp
               170                 175                 180
Lys Tyr Val Lys Val Phe His Lys Phe Leu Asn Ala Asn Arg His
               185                 190                 195
Lys Thr Arg His Thr Gly Glu Lys Pro Phe Lys Cys Lys Lys Cys
               200                 205                 210
Asp Glu Ser Phe Cys Met Leu Leu His Leu Ser Gln His Lys Arg
               215                 220                 225
Ile His Ile Arg Glu Asn Ser Tyr Gln Cys Glu Glu Cys Gly Lys
               230                 235                 240
Ala Phe Lys Trp Phe Ser Thr Leu Thr Arg His Lys Arg Ile His
               245                 250                 255
Thr Gly Glu Lys Pro Phe Lys Cys Glu Glu Cys Gly Lys Ala Phe
               260                 265                 270
Lys His Ser Ser Thr Leu Thr Thr His Lys Met Ile His Thr Gly
               275                 280                 285
Glu Lys Pro Tyr Arg Cys Glu Glu Cys Gly Lys Ala Phe Tyr His
               290                 295                 300
Ser Ser His Leu Thr Thr His Lys Val Ile His Thr Gly Glu Lys
               305                 310                 315
Pro Phe Lys Cys Glu Glu Cys Gly Lys Ala Phe Asn His Pro Ser
               320                 325                 330
Ala Leu Thr Thr His Lys Phe Ile His Val Lys Glu Lys Pro Tyr
               335                 340                 345
Lys Cys Glu Glu Cys Asp Lys Ala Phe Asn Arg Phe Ser Tyr Leu
               350                 355                 360
Thr Lys His Lys Ile Ile His Ser Gly Glu Lys Ser Tyr Lys Cys
               365                 370                 375
Glu Gln Cys Gly Lys Gly Phe Asn Trp Ser Ser Thr Leu Thr Lys
               380                 385                 390
His Arg Arg Ile His Thr Gly Glu Lys Pro Tyr Lys Cys Glu Glu
               395                 400                 405
Cys Gly Lys Ala Phe Asn Val Ser Ser His Leu Thr Thr His Lys
               410                 415                 420
Met Ile His Thr Gly Glu Lys Pro Tyr Lys Cys Glu Glu Cys Gly
               425                 430                 435
Lys Ala Phe Asn His Ser Ser Lys Leu Thr Ile His Lys Ile Ile
               440                 445                 450
His Thr Gly Glu Lys Pro Tyr Lys Cys Glu Glu Cys Gly Lys Ala
               455                 460                 465
Phe Asn Gln Ser Ser Asn Leu Thr Lys His Lys Ile Ile His Thr
               470                 475                 480
Gly Glu Lys Leu Tyr Lys Cys Glu Glu Cys Gly Lys Ala Phe Asn
               485                 490                 495
Arg Ser Ser Asn Leu Thr Thr His Lys Arg Ile His Thr Gly Glu
               500                 505                 510
Lys Pro Tyr Lys Cys Glu Glu Cys Gly Lys Ala Phe Asn Arg Ser
               515                 520                 525
Ser Asn Leu Thr Lys His Asn Ile Ile His Thr Gly Glu Lys Ser
```

```
                530                 535                 540
Tyr Lys Cys Glu Glu Cys Gly Lys Ala Phe Asn Gln Ser Ser Thr
                545                 550                 555
Leu Thr Lys His Arg Lys Ile Gln Gln Gly Met Val Ala His Ala
                560                 565                 570
Cys Asn Pro Asn Thr Leu Arg Gly Leu Gly Glu Gln Ile Ala Arg
                575                 580                 585
Ser Gly Val Gln Asp Gln Pro Gly Gln His Gly Lys Thr Pro Ser
                590                 595                 600
Leu Leu Lys Ile Gln Lys Phe Ala Gly Cys Gly Gly Arg Arg Leu
                605                 610                 615


<210> 35
<211> 418
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7021892CD1

<400> 35
Met Leu Leu Val Ala Thr Ile Leu Asp Ser Pro Gly Leu Asp Pro
  1                 5                  10                  15
Phe Tyr His Phe Asn Ser Met Ser Asn Asn Glu Arg Tyr Ser Ser
                 20                  25                  30
Phe Gln Leu Ala Cys Glu Ile Gln Lys Ser Ala Ala Met Thr Leu
                 35                  40                  45
His Val Cys Thr Arg Ile Ala Trp Tyr Lys Gly Tyr His Ile Val
                 50                  55                  60
Gly Lys Asn Leu Ser Asn Ser Asn Asn Leu Asn Asp Gly Arg Met
                 65                  70                  75
Lys Ser Glu Ser Asp Trp Ile Lys Lys Glu Gly Lys Gly Val Ala
                 80                  85                  90
Lys Val Gly Gly Asp Thr Leu Trp Tyr Lys Ser Pro Trp Gln Ala
                 95                 100                 105
Ala Leu Thr Pro Asp Leu Ser Cys Pro Gln Lys Gln Leu Glu Ala
                110                 115                 120
Arg Gly Glu Thr Pro Glu Gly Glu Thr Phe Ala Met Ala Glu His
                125                 130                 135
Phe Lys Gln Ile Ile Arg Cys Pro Val Cys Leu Lys Asp Leu Glu
                140                 145                 150
Glu Ala Val Gln Leu Lys Cys Gly Tyr Ala Cys Cys Leu Gln Cys
                155                 160                 165
Leu Asn Ser Leu Gln Lys Glu Pro Asp Gly Glu Gly Leu Leu Cys
                170                 175                 180
Arg Phe Cys Ser Val Val Ser Gln Lys Asp Asp Ile Lys Pro Lys
                185                 190                 195
Tyr Lys Leu Arg Ala Leu Val Ser Ile Ile Lys Glu Leu Glu Pro
                200                 205                 210
Lys Leu Lys Ser Val Leu Thr Met Asn Pro Arg Met Arg Lys Phe
                215                 220                 225
Gln Val Asp Met Thr Phe Asp Val Asp Thr Ala Asn Asn Tyr Leu
                230                 235                 240
Ile Ile Ser Glu Asp Leu Arg Ser Phe Arg Ser Gly Asp Leu Ser
                245                 250                 255
Gln Asn Arg Lys Glu Gln Ala Glu Arg Phe Asp Thr Ala Leu Cys
                260                 265                 270
Val Leu Gly Thr Pro Arg Phe Thr Ser Gly Arg His Tyr Trp Glu
                275                 280                 285
Val Asp Val Gly Thr Ser Gln Val Trp Asp Val Gly Val Cys Lys
                290                 295                 300
```

```
Glu Ser Val Asn Arg Gln Gly Lys Ile Val Leu Ser Ser Glu His
                305                 310                 315
Gly Phe Leu Thr Val Gly Cys Arg Glu Gly Lys Val Phe Ala Ala
                320                 325                 330
Ser Thr Val Pro Met Thr Pro Leu Trp Val Ser Pro Gln Leu His
                335                 340                 345
Arg Val Gly Ile Phe Leu Asp Val Gly Met Arg Ser Ile Ala Phe
                350                 355                 360
Tyr Asn Val Ser Asp Gly Cys His Ile Tyr Thr Phe Ile Glu Ile
                365                 370                 375
Pro Val Cys Glu Pro Trp Arg Pro Phe Phe Ala His Lys Arg Gly
                380                 385                 390
Ser Gln Asp Asp Gln Ser Ile Leu Ser Ile Cys Ser Val Ile Asn
                395                 400                 405
Pro Ser Ala Ala Ser Ala Pro Val Ser Ser Glu Gly Lys
                410                 415
```

<210> 36
<211> 1010
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7492673CB1

<400> 36
atggcggatg acgccagtgc agagggcggg ggcgggcggg cgcgggcggg cgggcgggtg 60
gtgtgggcgc cggaggcccc gggggccctg ggaaggggaa ctgatgtgac ttccgctgag 120
ggtttcagca gtggcatccg ggtccggggt cacggccgtg gatggggccg gggccggggt 180
cgcggccgtg gatggggccg gggccaagac cgcggagttc gcggaggcaa ggcccaggat 240
aaggagcgga tgcccgtcac caagctggac ggcctggcca aggacatgaa gatcaagtcc 300
ctagaggaga tttatctctt ctcgctgccc atcaaggaat ctgagatcat tgacttttc 360
ctgggggcct ctctcaagga cgaggttttg aagattatgc ctgtgcggaa ggagacccgc 420
gccggccagc gcaccaggtt caaggcgttt gttgccatca gggactacaa tggctacgct 480
ggtgggttga agtgctccaa ggaggtggcc gccgccatcc gcggggccat catcctgacc 540
aagctctcca ttgtcaccgt gcgcagaggc tactgggggа acaagatcgg caagctccac 600
accgtccctt gcaaggtgac aggccgctgc ggctctgtgc tggtgccctt catccccgcg 660
cccaggggca ctggcatcgt ctcagctcct gtgcccaaga agctgctcat gatagctggt 720
atctacgact gctacacctc agccaggggc tgcactgcca ccctgggcaa cttcgccaac 780
gccaccttgg atgctgtctc taacatctac agctacctgc ccctgaagga cttctggaag 840
gagactgtat tcaccaagtc tccctatcag gaattcactg accacctcgc caagacccac 900
accagagtct ccgtgcagag gacccaggct ccagctgtga ctacaacata gggtttttat 960
gcaagaaaaa taaagtgaat taaagcctat tactgtcaaa aaaaaaaaaa 1010

<210> 37
<211> 612
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7990930CB1

<400> 37
taagaagtga tgggagctgc atctgctgca agaatgaagg ctgttctcag taatcagact 60
gtccacattc cagaaaatgt cgacatcact ctgaaggggc acacagttct tgtgaaaggc 120
cctagaggaa ccctgcagag ggacttcagt cacgtcagta tagaactcgg gctccttgga 180
aagaaaaagc agaggctcca gattgacaaa tgcatgctga gaagatggga attggctgcc 240
atgcgtgcta tctgtagtca tgcacagaag atgatcaagg gcgttatact gggcttctgt 300
gacaagatga ggtctctgtg ttctcacttc cccatcaatg tcattatgca ggagaatagc 360
tctattgtag aagtcagaac tttcttgagc gaaaaatata tctgcagggt tcgggtgagc 420
ccaggtgttg cttgttcaat atcccaagcc cagaaagatg agttcatcct taaaggaaat 480

gacattgaac ttgtttcaaa ttcagctgct ttgactcagc aaaccacaac agctaaaaac 540
caggatatca gaacattttt ggatggtatc tatgtctctg aaaaagggac agttcagcag 600
gctgatgagt aa 612

<210> 38
<211> 2663
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7037554CB1

<400> 38
ctgagcccta acgagaaggc tgggcctagg ccgctggatg ctggagtgaa aggaagggag 60
aaagggaaaa agcgggaaga gtcgagaagg ctgagtgtta agaggccaag tgcgacgcgc 120
gtatccgggc agacggactg acggacgggc ccgtgcttct gccgcggctg cggcgcccgc 180
gcgagtcgcg tctaagcggc ggcggcggtg gcagcggcgg aaaccgaagg ggagccatgg 240
cggctgacag tcgggaggag aaagatggag aacttaatgt tctggatgat attttaactg 300
aagtaccaga acaagatgat gaactgtata atccagagag tgaacaagat aaaaatgaga 360
aaaagggatc aaaaagaaaa agtgatcgaa tggaatctac tgataccaaa cgacaaaagc 420
cttctgtcca ttcaagacaa ctggtttcta agccactgag ctcatctgtt agcaataaca 480
aaagaatagt tagtacaaaa ggaaagtcag ccacagagta taaaaatgag gaatatcaaa 540
gatctgaaag aaacaagcgt ctagatgctg atcggaaaat tcgtctatca agtagtgcct 600
ccagagaacc ttataagaat caacctgaaa aaacctgtgt ccggaaaagg gatcctgaaa 660
ggagggccaa atctcctacg ccagatggtt ctgagagaat tgggcttgaa gtggatagac 720
gtgcaagcag atccagccag tcttctaagg aagaagtgaa ctctgaagaa tatggctctg 780
accatgagac tggcagcagt ggttcttctg atgagcaagg gaacaacact gagaatgagg 840
aggaaggagt ggaagaagat gtggaggaag atgaagaagt agaagaagat gcagaagaag 900
atgaagaggt ggatgaagat ggagaggagg aggaggaaga ggaggaggag gaagaggagg 960
aggaggagga ggaagaagaa gaatatgaac aggatgagag agaccagaaa gaggagggaa 1020
atgattatga cactcgaagt gaggccagtg actctggttc tgaatctgtt tccttcacag 1080
atgggtctgt cagatctggt tcaggcacag atggatcaga tgagaaaaag aaggaaagga 1140
agagagctag aggcatatct ccaattgttt ttgatagaag tggaagctct gcatcagagt 1200
catatgcagg ttcagaaaag aagcatgaga aattatcatc ttccgttcgt gctgtccgaa 1260
aagatcaaac cagtaaactc aaatatgtgc ttcaagatgc aagatttttc ctcataaaga 1320
gtaacaacca tgagaatgtg tctcttgcca aagcgaaggg tgtatggtcc acgctccctg 1380
taaatgagaa gaaattaaat cttgcattta gatctgcaag gagtgttatc ttaatatttt 1440
ctgtcagaga gagtggaaaa tttcaagggt ttgcaagact ttcttcagaa tcacatcacg 1500
gaggatctcc tatacactgg gtgcttccag caggaatgag tgctaaaatg ctgggaggtg 1560
tctttaaaat tgactggatt tgcaggcgtg aattaccctt cactaagtcg gctcatctca 1620
ccaatccttg gaatgaacat aaaccagtaa agatcggacg tgatggacag gaaattgaac 1680
ttgaatgtgg aacccagctt tgtcttctgt ttccccccga tgaaagtatt gacttgtatc 1740
aggtcattca taaaatgcgt cacaagagaa gaatgcattc tcagccccga tcacgaggac 1800
gtccatcccg tcgagaacca gtccgggatg tgggaaggcg tcgaccagaa gattatgata 1860
ttcataacag cagaaagaaa ccaaggattg actatccccc tgagtttcac cagagaccag 1920
ggtatttaaa ggatccacga taccaggaag tggacagttt cacaaatctt attcccaaca 1980
gacgattttc aggagttcgc cgagatgtgt ttttaaatgg gtcctacaat gattatgtga 2040
gggaatttca taacatggga ccaccaccac cttggcaagg aatgcccctt tacccaggaa 2100
tggaacaacc tccacaccat ccttactatc agcaccatgc tccacctcct caagctcatc 2160
ccccttactc aggacatcat ccagtaccac atgaagcaag atacagagat aaacgagtac 2220
atgattatga tatgagggtg gatgatttcc ttcgtcgcac acaagctgtt gtcagtggcc 2280
ggagaagtag accccgtgaa agagaccggg aacgagagcg agaccgccct agagataaca 2340
gacgagacag agagcgagat agaggacgtg atagagaaag agaaagagag cgcttatgtg 2400
atcgagacag agaccgaggg gagagaggtc gatatagaag ataatgggct tttggaagca 2460
ctgattgttt aaagatacaa aaaatcttgt attttttttt ggtgtgtgtt tacaagtagt 2520
aaatttattt tcagctgtct gcctatgaag ttcattgtgt agaaggattt attatgaccc 2580
cctttgttcc aagcatgcag tatcataaga actggaaaaa ctcaaatccg ccagaaatcc 2640
acagctgaca gttgaattga cat 2663

<210> 39
<211> 7188
<212> DNA

<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1515347CB1

<400> 39
tagcgccttt ctggccttgg atcacccggc gcgtcccaac cgcaccttga cccacaagtg 60
gcttccggtc cgctgctgtg aatgcctcgc tttaggagaa ccctgacggt ctccaaacca 120
gcttctccca ttggagggcc gacccaagga ggaaaagacc agactgcttg aaagagcgcc 180
tggatcagat ttatttagtc taacgagcgg cgctgttctc aagctccagt ctatggcaga 240
gacttgctaa ggatttgtgc cctgcctagc catggaaggg tacagtggcg tgggtccctg 300
gatgctcgtc gtgggaagga ggccgggcca gcgcacagtt acacttcatc ctcagaaagt 360
ccaagtgagc tgatgttgac gctttgtcgg tgtggagagt ctctgcagga tgttattgac 420
agggtggcct ttgtgattcc tccggtggtg gcagcacccc cgtccctacg ggtgccgcgg 480
ccgccacccc tgtacagcca cagaatgagg atcttgaggc agggcctgag agagcacgct 540
gcgccgtact tccagcagct gcggcagacc acggctccac gcctgctgca gttccctgag 600
ctgaggctgg tgcagttcga ctcagggaag ttggaagctt tagctatctt gcttcagaaa 660
ttgaaatctg aaggacgtcg ggtgctgatt ttatcacaga tgattcttat gttggacatt 720
ttagagatgt tcttgaactt ccattacctc acctatgtaa gaatcgatga aaatgccagc 780
agtgagcaac ggcaggaact gatgaggagt ttcaacagag acaggcggat tttttgtgcc 840
attctctcca ctcacagccg taccacaggt ataaaccttg tagaggcgga caccgtcgtg 900
ttttatgaca atgacctgaa tccagtgatg gatgccaaag ctcaggagtg gtgcgatagg 960
atcgggagat gcaaagacat ccacatatac aggcttgtga gtggcaattc cattgaagag 1020
aaattgttga aaaatggaac taaagatctg atccgagaag tggctgctca gggaaatgac 1080
tactccatgg ctttcttaac tcagcgaacc atccaggagc tgtttgaagt ttattctccc 1140
atggatgatg ctggcttccc ggtcaaagct gaggagtttg tggtgctttc tcaggaacct 1200
tctgtcacgg aaaccattgc acccaaaatt gcaagacctt tcatagaggc cctcaagagt 1260
attgagtatc tggaggagga tgcccagaag tccgcacagg agggggtgct gggaccacac 1320
actgatgctc tgtcatcaga ctctgagaac atgccgtgtg atgaagaacc atcccaatta 1380
gaggagctag ctgacttcat ggagcagctt acaccaattg aaaaatatgc tttaaattac 1440
ctggaattat tccatacttc tattgagcaa gaaaaggaga gaaacagtga ggacgcagtg 1500
atgactgcag tgagggcatg ggagttctgg aacctgaaga ccctgcagga gagggaggcc 1560
cggctgcggc tggagcagga ggaggcggag ctcctgacct acacgcgaga ggatgcctac 1620
agcatggagt atgtctacga agatgtcgat gggcagacag aagtcatgcc gctctggacc 1680
ccacccaccc cgccgcagga cgacagcgac atctacctcg actcggtcat gtgtctcatg 1740
tatgaagcca ctcccatccc agaggctaag ctgccccctg tgtacgtgag gaaggagcgg 1800
aagcgacaca aaacagaccc ctcagctgca ggcaggaaga agaagcagcg tcacggggag 1860
gcggtcgtcc ctcctcggtc cctgtttgac cgcgcaacac caggacttct gaaaattcgc 1920
agagagggca aggagcagaa gaagaatatt ctgctgaagc agcaggtgcc attcgccaag 1980
cccctgccaa cttttgccaa acccacagct gagcctggtc aagacaaccc cgagtggctc 2040
atcagtgagg actgggcgct gctgcaggct gtaaagcagt tactggagct gcctttgaac 2100
ctcacaatcg tgtcacctgc tcacacacct aattgggatc ttgtcagtga cgttgttaac 2160
tcctgtagcc gaatctaccg ctcttccaaa cagtgccgga atcgctacga gaatgtcatc 2220
attccacgag aggaggggaa gagtaaaaac aaccgtcctc tccgtacgag ccagatctat 2280
gcccaggatg agaatgccac acacacccag ctgtacacga gccactttga cttaatgaaa 2340
atgactgctg gcaagaggag tcccccaatc aaacctctgc ttggcatgaa tccctttcag 2400
aagaacccca agcacgcgtc tgtgttggca gaaagtggaa tcaactatga caagccgctg 2460
cctcccatcc aggtggcatc tctccgtgca gagcgaatcg caaaagagaa aaaggctctg 2520
gctgatcagc agaaggcaca gcagccggcc gtggcccagc caccccgcc ccagccgcag 2580
cccccaccac ccccgcagca gccaccgcca ccgctgccac aaccacaggc agcgggcagc 2640
cagccgccag cagggccacc agctgtccag ccccaacccc agccacagcc ccagacccag 2700
ccacagcctg tgcaggcccc agcgaaggcg cagcccgcaa tcacgacggg gggcagtgca 2760
gccgtactgg caggaaccat taaaacatca gttactggga cgagcatgcc cactggtgcc 2820
gtgagtggaa atgtgatcgt gaacaccatc gcaggggtcc cagctgccac cttccagtcc 2880
atcaacaagc gcctggcgtc gccagtggct cctggggcct tgactacgcc gggaggctct 2940
gctcccgccc aggtggtgca cacccagccc ccgccacggg cagtcggctc cccagccacg 3000
gcgacccctg acctggtgtc catggcaacg actcagggtg ttcgagcggt cacttctgtg 3060
acagcctcgg ccgtggtcac taccaacctg accccagtgc agaccccggc acggtctttg 3120
gtgccccaag tgtcccaagc cacaggagtt cagctccctg gaaaaaccat cacacctgca 3180
catttccagc ttctcaggca gcagcagcag cagcagcaac aacagcagca gcagcagcag 3240
cagcagcagc agcagcagca gcagcagcaa cagcagcagc agcaacagac gacgacgacc 3300
tctcaggtgc aagttccaca gatccagggc caggcccagt ccccagcaca gatcaaagct 3360

gtgggcaagc tgacgccgga acacctcatc aaaatgcaga agcagaaact gcagatgccc 3420
ccgcagcccc caccgccaca ggcccagtct gcgcccccgc agccaacagc ccaagtgcaa 3480
gtgcagacct cgcagccgcc gcagcagcag agcccccagc tcacgacggt cacggcccca 3540
aggcctggtg ccctgctgac gggcaccacc gtggccaacc tccaggtggc ccggctcctg 3600
caggcgcaag ggcagatgca gacccaggca ccccagccag cccaggtggc cttggcgaag 3660
cctccggtgg tgtccgtccc ggcagctgtg gtctcctcac cgggagtcac caccctgccc 3720
atgaacgtcg cggggatcag cgtggcgatc ggtcagccac agaaggcagc aggacagacc 3780
gtggtggccc aggcccgtca catgcagcag ctgctgaagc tgaagcagca ggccgtccag 3840
cagcagaagg ccatccagcc ccaggctgca cagggcccgg cagccgtcca gcagaagatc 3900
accgcacagc agatcaccac ccctggcgcg cagcagaagg ttgcctacgc cgcgcagccg 3960
gcccttaaga cccagtttct taccacaccc atctcccagg cccagaaact ggccggggcc 4020
cagcaagtgc agacccagat ccaggttgca aaacttcctc aagttgttca acagcaaaca 4080
cccgtggcca gcatccagca agttgcctct gcttcccagc aggcttctcc acagactgtg 4140
gcgctcacgc aggcgacggc ggccgggcag caggtgcaga tgatccctgc agtgaccgcg 4200
actgcccagg tggttcagca gaaactcatt cagcagcagg tggtgaccac ggcgtcggcc 4260
ccgctccaga ctccaggcgc tcccaaccca gcccaggtgc ccgccagctc cgacagccca 4320
agccagcagc ccaagttaca gatgagggtc cctgctgtca ggctaaagac acctactaag 4380
cctccgtgcc agtagtcagg gcagcagggc tgcctctcat ctaaagcaaa actaccttcc 4440
tcacagaaaa cgctttatta gtgaaccttg ggaccatgtc acgcaagaga ttcagcactg 4500
ggaaagatat aattgaaaca aaatagtgta atcattttat taaaatgcat cccacactgc 4560
aggacaaatg gtccttatgg agtgccgcgt tctctgtact acgtggctca tggaaaaagt 4620
gacaacatgg cttcctctaa atcatttcac ctttcagtcc ccacccgcac ccgtccccta 4680
gagccatagt actgtgttct gaaagccatt tagaatttct ttgtgagcat gtagtgcttt 4740
gcacgccaca gaagccgtct gccgtgtgtg aggagcatac aatggacttt ctaaagataa 4800
ggcgtgggct tccacagtgt ctgccagagt ttagttcttt ataccttact gaaaaatgcc 4860
tcgtggtctt cgcagagggg aaggcctgtc taaagtcaat catccgagat gggttttcca 4920
ttccaaagaa aggcaatatg gttccttcct tccctcctaa aatatgactt aacttttaag 4980
agaaatgttc tgacacccac ctaaacacac aaggcacgtt cctggcctgt gttcaaggga 5040
aatgatcagt cattgcattg ttattccaaa gagcagccaa cagtggcctc ccccaggccc 5100
taccctgcaa tgggattcgc tttcatttaa tggaaacttc tgggactgat gcccaactca 5160
gtgcactcaa gacgcatctc cagttttcgg gggaagctgg tatttgacat agtgtgttaa 5220
acagctcctg agaacctttg ggacactctg ccatggctgg cgtgaggccc agaggaccac 5280
gcagaggcaa tggtagtaca gatgtcacag ctgagggtac gatgaggcct gggctcagtg 5340
agccaggacg aatgtgacag acacccccttg ctgccacagt cagcccttg acgaaggtgg 5400
gctggtgatt ctggaagtat tggctatagc ggtgggccca gtcaactctt ccttgtggac 5460
ttacgacagc agattttctc taggataagc ttgtgtggtt ctgccagtga agcagagaac 5520
cacctgtgct gttgtggaag gcgtgccgtt gaggggggaaa acgaagccca gtatttgcta 5580
ctgtttttcc tttttttact atgacaggaa aataaatgca attttagtgg aattgattga 5640
cagtgtctcc ttactttgaa gttttcacca aagcaaaaag gtccatatcc aatagtatcc 5700
tttgtgctgt ggcttgattt tggcctattt tacattattt ggtccaggaa attaggttat 5760
attaggtttt ttgtatacta aaaatcagtt atggcacaat aaagattttc tgtttttaaa 5820
ttgtatttca tctgcttcct ccccattctc tcactttaag tgacattgag gaaggtattc 5880
tgtcccacag gtttctgtgg acagcgatac agcaggagtc agtgaaatca actggggagc 5940
tcacttgagc tcttgataag aaatgtggag aaaagtaaaa accaagcttt gaagaaacag 6000
aagaaattaa tcttttagtt agttgaacat accaaagcag aggactggaa tctgtttgtt 6060
ctaaccaacc cgttctccct ggcttggcac gtgccgtgag agcgcagctt gccggaggga 6120
gggccgctgt gtgcgcctca catctggctc ccagtggaaa cttttactcc tcctcatccg 6180
cagatgtgat agaactgaag tatctaggaa ttctgccttt gtcatttgtt ttaatttgtg 6240
tgccctgttc attttttttg tctttcccaa atcttggtag tctccttata gttgaagata 6300
aaatgttgag tgcacttatt ttagaatatc ctagacataa ctgtctaagt aaaagcgctc 6360
tattaatcta aaacactaca agagaattta acaccatctc tcaaatgctt ttttggagag 6420
cttaatggga ttctgaatat ttgcaatgtg gagtttccgc cccgatctca cgtcagtgag 6480
ggtctcctgt ctctcaagtg tgtttccttt ggctgttccc taatacaaaa cacggacata 6540
tttttactcg tagcactcaa tttagtaact tctagatgct accgttgacc tgagttaaat 6600
tcatttagtc gtgtacgtaa aaactctcct tttagtgtgt tattttcttg gccttccctt 6660
ttaaaggtta aagtttctaa cctaagaatt aagtacgcgt tcaggaagct gttgtctagg 6720
ccttcccctt gtgaatctgg gttcattcca atacggcaag taagagttgg aaactttgag 6780
aacacagact ataaaggcag cagcccgaac actgtcagac tctaattggc gaccctggga 6840
aacagttgcc ctgctattct ttaaagaaag acgtttattc tgatgataaa aacagttagc 6900
cagactgttt ttaaagcacc tggcgggaag cagaaggttg gatccaagcc cttgttcaga 6960
tttggtgcct gataagacag gggtttctct ttttgtgacc tttattatta ttattttgtt 7020
aactgttgta accagttagc tgttgtgttt taagatagaa aggaacaaga ctaaaattgt 7080
aaatactttg taaacatcag catttgtact tgaatagtag gattttaaag ggcattgata 7140

gcataccaaa caaaaggcaa aataaagtga ccttttttata tatttttt 7188

<210> 40
<211> 1972
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3464492CB1

<400> 40
tatgtatctt agaaaagcta aactctttat agctgtcctt tcttaaccta ttttacaata 60
tttaaaacac tggaaagatg aatttatatg tcgtattaat tgcaaatgtc cttttaattc 120
aaggttgagt taaaagaata tctgaaaata aatgatacaa tatatgaagt tgacagcaaa 180
gctgagaatg gcatgacttt ttcacgtctt cttaattata agtattctga taccctaaaa 240
aagatggatc ctgatcactt ggtagcattg gtgacagaag ttattcccaa ttattcctgc 300
ttagtttttt gtcctagtaa gaagaactgt gaaaatgtag cagaaatgat atgcaaattt 360
ttaagcaagg aatatctgaa acataaggag aaagaaaaat gtgaggtgat taagaacttg 420
aagaatattg gcaatggcaa cctgtgtcct gttttaaagc gcactatccc atttggagtt 480
gcctatcacc acagtggctt aacaagtgat gaaaggaaac tcttggagga ggcctactcc 540
acaggagtgc tctgtctttt tacctgcaca tctaccctag cggcaggtgt caacctacca 600
gctcgaagag ttattttaag agctccctat gttgctaagg aatttttaaa gaggaatcaa 660
tataaacaga tgattggcag agctggtcgt gctggaatag atactattgg ggagagtatc 720
ctcatattgc aagaaaaaga caaacaacag gtattggagt taataactaa accattggaa 780
aactgttaca gccatcttgt tcaggaattc accaagggaa tccaaacatt atttctctct 840
ttgattggtt tgaagattgc aacgaatctt gatgacatct atcatttcat gaatggtaca 900
ttttttggtg ttcagcaaaa ggttttattg aaagaaaaaa gtctctggga aataactgtt 960
gaatcactta gatacctgac agaaaaagga ctcctacaaa aagacactat ttataagtct 1020
gaagaagagg tccaatataa ttttcatatt acaaagttgg gacgtgcttc atttaaggga 1080
actatagatt tagcttattg tgacattctg tacagagact tgaagaaagg tcttgaagga 1140
cttgtgcttg aaagccttct tcatctaatc tacctaacaa ccccctatga tctggtttca 1200
cagtgtaacc ctgattggat gatatacttc aggcagttta gccaactcag tccagcagaa 1260
caaaatgtag ctgccattct tggagtctct gaaagcttta ttgggaagaa agcatcaggc 1320
caagccatcg gaaagaaggt ggacaagaac gttgtcaaca ggctatatct gtcttttgtt 1380
ctttatacct tgctcaaaga gaccaacatt tggactgtat ctgaaaaatt taatatgcct 1440
cgaggatata tacaaaatct tctcactgga actgcctcat tctcatcttg tgtgttacat 1500
ttctgtgagg agcttgagga gttttgggtt tacagagccc ttttggtaga acttaccaag 1560
aagctgactt actgtgtaaa ggcagaatta atccctctca tggaagttac tggagtttta 1620
gagggtcgag caaaacagtt atacagtgca ggttacaaaa gtctaatgca cttagctaat 1680
gcaaatcctg aagtgctcgt aaggacaatt gatcatttat caagacgcca agccaagcaa 1740
attgtttcat cagcaaagat gctgttgcat gaaaaagcag aagccctgca agaagaggta 1800
gaagagttac taagattgcc ttctgatttc ctggtgctgt ggcttcttcc actgacaaag 1860
catgaagcta tctgatgaga attttcatat atgaattatt tattgtacat gtgtgttatt 1920
aaagataact tatactttat aaatgaaaaa atacatttca gttatacatt ag 1972

<210> 41
<211> 1857
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1794336CB1

<400> 41
atggaagaat ttaaaagcca tagtcctgag agatcaattt tcagcgctat ctgggaaggc 60
aactgtcatt ttgagcaaca tcagggacaa gaggagggtt attttagaca acttatgatt 120
aaccatgaaa acatgcccat ttttagccaa catactttac tcactcaaga attttatgat 180
agagagaaaa tctctgaatg taaaaagtgt agaaaaatct tcagttacca tttatttttt 240
agtcaccaca aaagaactca ttctaaagaa ctttctgaat gtaaagaatg cacagaaatt 300
gttaatacac catgcctttt taaacaacag actattcaaa atggtgacaa atgcaatgag 360
tgtaaagaat gttggaaggc ctttgttcat tgctcacact ttaaacatct aagaattcat 420

aatggtgaaa aacgctatga atgtaacgaa tgtgggaagg cctttaatta tggctcagaa 480
cttactctac atcaaagaat tcacactggt gagaaacctt atgaatgtaa agaatgtggg 540
aaggccttta gacagcgatc acagcttact caacatcaga gacttcatac tggtgaaaaa 600
ccctatgaat gtaagcaatg tgggaaggct tttattcgtg gctttcaact tactgaacac 660
ctgcgactcc atactggaga gaaaccttat gaatgtaaag aatgtggaaa gacttttagg 720
catcgctcac atcttactat acatcagaga attcatactg gtgagaaacc ctatgaatgt 780
cgggaatgtg ggaaggcctt tagctatcac tcaagcttct cacaccatca gaaaattcat 840
tctggcaaga aaccttatga atgtcatgaa tgtgggaagg ctttttgtga tggcttacaa 900
ctaaccctac atcagaggat tcatactggt gagaaaccct atgagtgtaa ggaatgtggg 960
aagactttta gacagtgttc acacctcaaa agacatcaga gaattcatac tggtgagaaa 1020
cctcatgaat gcatgatatg tggtaaggcc tttagacttc attcacacct tattcaacat 1080
caaagaattc atactggtga gaaaccctat gaatgtaagg aatgtgggaa ggcctttagc 1140
tatcattcaa gcttctcaca ccatcagaga attcattctg gaaagaaacc ttatcaatgc 1200
gggaaggcgt ttaatcatag attacaactt aacttacatc agactcttca tactggcgag 1260
aagccagtca ggtttcctct cctccctccc catcctagtc tagcatcatg aaatttgttg 1320
atggaaatat acattgtctt tgttttattt aaggaaaaag ttggtaagcc accagtttca 1380
ttagtattta tttatttaaa taatccttga aaatcaaatg tttttaagtg agaaagtgga 1440
taactaaagt gtaaattttg gtctcatttt aatgctgttt ctgagttatt gtattaacca 1500
ttattgagaa taaatttttt tattgagaat aaattttcta aaattacatg tttcactgaa 1560
accagtgaaa tatttcactg aaacagtgaa ataaaatgaa aacaatttcc tcccttccct 1620
caaatctctg agataaccag tatgtgtatt tcttcttcct ttatgctcaa acaacctgta 1680
cacagaaagc atctttcgca tgtgtaggga ggagctttta ccccccatac agacctttga 1740
caggcgtcct tgtccctttt cataaataat cttgaatact tttccaggtc agtacgttca 1800
agaatgtaac ccattctttg taagggtcca gaattttcct agaatggaat gtttggc 1857

<210> 42
<211> 2454
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2903694CB1

<400> 42
atggatcgag atttagaaca ggctctggat cgcacagaga atatcactga aattgcccaa 60
cagagacgtc ctagaaggag atactcacct agggcgggaa aaactctgca ggaaaaactt 120
tatgacattt atgttgaaga atgtggaaaa gagcctgagg atcctcagga attgagaagc 180
aatgtaaact tgttagaaaa gcttgttagg agagagtcct tgccatgttt actggtcaat 240
ctatacccag gcaatcaggg gtattctgtg atgctccaga gagaagatgg gtcctttgca 300
gagaccattc ggctgcctta tgaagaaagg gcattgctgg actacttgga tgcagaagaa 360
ttacccccctg ctttgggtga tgtcctggat aaagcttcgg ttaacatttt tcatagtggg 420
tgtgtcatag tagaagttcg tgactacagg cagtccagta atatgcaacc tcctggttac 480
caaagcaggc atattcttct acgtccaacc atgcagactt tagcccctga ggtgaagacg 540
atgacaagag atggcgagaa atggagccag gaagacaaat ttcctcttga gagtcaactg 600
atcttagcga cagctgaacc actgtgtctt gatccttctg tagcagttgc ctgcactgca 660
aacaggctgc tgtacaacaa gcaaaagatg aataccgacc cgatggaaca gtgcctccag 720
aggtattcgt ggccctctgt aaagccacag caggagcagt ctgactgtcc acctcctcct 780
gagctgagag tgtcgacttc tggccaaaaa gaagaaagaa aagtaggtca gccttgtgag 840
ctgaacatta ctaaagcggg aagttgtgta gacacgtgga aaggcagacc ctgtgatttg 900
gccgtgcctt cagaagtgga tgtggagaaa cttgctaaag ggtatcagtc cgtcacagct 960
gctgacccac agctcccagt ctggccagcc caggaggtag aagacccttt cagacatgca 1020
tgggaagctg gctgtcaggc ctgggacacc aagccaaaca tcatgcagtc gtttaatgat 1080
ccgcttctct gtggtaaaat acggccacgt aaaaaagcca ggcagaagag ccagaagtct 1140
ccctggcagc ccttcccaga tgaccattca gcttgtctca ggcctgggtc agagactgat 1200
gctgggaggg cagtgagtca ggcccaggaa tcggtgcaga gcaaagtcaa aggtccaggc 1260
aagatgtcac acagctccag tggcccagcc agtgtcagtc agctctcttc atggaaaaca 1320
ccagaacagc ctgatcctgt gtgggtccag tcttcagtat ccgggaaggg agagaaacat 1380
ccacctcccc gcacccaact tccctcaagc tcaggaaaga tttcctcagg taacagtttt 1440
cccccacaac aggcaggcag ccctcttaag cgtccatttc ctgctgctgc tcctgctgta 1500
gctgctgctg ctcctgctcc tgctcctgct cctgctgctg ctcctgcttt agctgctgct 1560
gctgtggcgg cggcggccgg tggggcggca ccaagccatt ctcagaagcc ctctgtgcct 1620
cttattaaag ctagcaggcg ccgtccagct gccgggcgcc ccaccagatt cgtaaaaata 1680

gccccagcca ttcaggtccg gacaggctcc actggcctaa aggccaccaa cgtggagggc 1740
ccagtccggg gagcccaggt tttggggtgc agtttcaagc ctgtgcaggc ccctggctcg 1800
ggtgcccccg ctcctgcagg aatcagtggc agtggccttc agtcctcagg aggtccacta 1860
ccagatgcaa ggcccggtgc agtgcaggca tcttctccag cacccccttca gtttttccta 1920
aatactccgg aaggtctcag gcctctgaca ctccaggttc cgcagggctg ggcggttctg 1980
accggcccgc agcagcagtc ccatcagctg gtttccctgc agcagctcca gcagcccaca 2040
gctgctcacc ctcctcagcc agggccacag ggttccacac taggtttgag cacgcaaggg 2100
caggccttcc ctgctcagca acttcttaat gtgaacctca ctggagcagg tagtggtctg 2160
cagccccagc cccaggctgc tgtgttgagt ctgcttggct ctgcccaggt tcctcagcag 2220
ggtgtccagc tccccttttgt cttggggcag cagccacagc cgctgctgct gctgcagcca 2280
caaccacagc cacagcagat ccagctccag acacagcctt tgagagtcct gcagcagcca 2340
gtgtttttgg caacaggcgc tgttcagata gtgcagccac atccaggtgt gcaagcaggg 2400
agccagttgg taggtcagag gaagggaggc aagccaaccc ctccagctcc ctga 2454

<210> 43
<211> 4409
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6975426CB1

<400> 43
agttctggga gtcaagcttt tactaaactt gggtgactgg tcctaagtgg cgtacatcat 60
actggagtga gaattgactt gcatgctatt gcttttggtt tctgcaggtt ttgtagtgca 120
tccctgagca tgagcacaga atgaagcggc gtttggatga ccaggagtca ccggtgtatg 180
cagcccagca gcgtcggatc cctggcagca cagaggcttt tcctcaccag caccgggtgc 240
ttgcccctgc ccctcctgtg tatgaagcag tgtctgagac catgcagtca gctacgggaa 300
ttcagtactc tgtaacaccc agctaccagg tttcagccat gccacagagc tccggcagtc 360
atgggcccgc tatagcagca gttcatagca gccatcatca cccaacagcg gtgcagcccc 420
acggaggcca ggtggtccag agtcatgctc atccagcccc accagttgca ccagtgcagg 480
gacagcagca atttcagagg ctgaaggtgg aggatgcgct atcttatctt gaccaggtga 540
agctgcagtt tggtagtcag cctcaggtct acaatgattt ccttgacatc atgaaggaat 600
ttaaatctca gagcatcgac accccaggag tgattagtcg tgtgtcccag ctattcaaag 660
gccacccga tctgataatg ggattcaaca ccttcttgcc ccctggctac aaaattgagg 720
tgcaaaccaa tgacatggtg aatgtgacaa ctcctggcca ggttcatcag attcccaccc 780
atggcatcca gccacagcct caaccaccac cccaacatcc ttcccagcct tcagcccagt 840
cagccccagc tcctgcccag ccagctcctc agcccccacc tgccaaagtc agcaagccct 900
cccaactgca agcacatact ccggccagtc agcagactcc cccacttcca ccgtatgcat 960
ccccacgttc tccgccggtc cagcctcaca caccagtgac aatctcgttg ggaacggccc 1020
catccttgca gaacaatcaa cctgtggagt ttaatcatgc catcaactat gttaataaga 1080
tcaagaacag atttcagggc caaccagaca tctacaaagc attcctggag attttgcaca 1140
catatcagaa agagcagaga aatgccaagg aagctggagg aaactacact ccagcattga 1200
cagagcagga ggtgtatgcc caggttgctc gtctctttaa aaaccaggaa gatttgttgt 1260
cagagtttgg acaattccta ccagatgcca acagctccgt gcttttaagc aaaacaactg 1320
ctgagaaggt tgattctgtg agaaatgatc atggaggcac tgtcaagaag ccccaactga 1380
acaacaagcc gcagaggccc agccagaatg gctgccagat ccgcagacat cctacaggaa 1440
ccacccctcc agttaagaag aaacccaaac tgctcaatct gaaggattct tctatggcag 1500
atgccagcaa acatggtggt ggaacagaat cgttatttttt tgataaggtc cgaaaggctc 1560
ttcggagtgc agaagcctac gaaaatttcc tacgctgtct tgttatttttt aaccaggagg 1620
tgatctctcg tgctgagctt gtgcaactag tctctccttt cctggggaaa ttccctgagt 1680
tgtttaattg gtttaaaaac tttctgggct ataaggagtc tgtacatctg gaaacttatc 1740
caaaggagcg agccacagag ggcattgcta tggagataga ttatgcttct tgtaaacgat 1800
tgggctccag ctatcgagcc ttaccaaaga gttaccagca gcccaagtgt acaggacgga 1860
ctcctctctg taaagaggtt ttaaatgata cctgggtttc cttcccttcg tggtctgagg 1920
actctacctt tgtgagttcc aagaagactc aatatgaaga acatatttat cgttgtgaag 1980
atgaacgctt tgagcttgat gtagttttag agaccaatct ggcaacaatc cgggttctgg 2040
aagcaataca gaagaagctt tcccgcttgt ctgctgaaga acaagccaaa tttcgcttgg 2100
acaacaccct tgggggcaca tcagaagtca tccatagaaa agcactccag aggatatatg 2160
ctgataaagc agctgacatc attgatggtc tgagaaagaa tccctccatt gctgttccaa 2220
ttgtccttaa aaggttgaag atgaaagagg aagaatggcg agaagctcag agaggcttta 2280
acaaagtatg gcgagaacaa aatgagaaat actacttgaa gtctctggac caccagggga 2340

tcaactttaa acagaatgac accaaggtcc tgaggtctaa gagcttactc aatgagattg 2400
agagtatcta tgatgagagg caagagcagg ctacggagga gaatgctggt gtacctgttg 2460
gcccacacct ctcacttgcg tatgaagaca aacaaatact ggaagatgct gctgctctga 2520
ttatccacca tgtgaagagg cagacaggca ttcagaagga ggacaaatat aagataaaac 2580
aaatcatgca tcattttatt ccagatttgc tctttgccca aagaggtgat ctctcagatg 2640
tggaggaaga ggaagaagaa gagatggatg tagatgaagc cacaggggca gttaagaagc 2700
acaatggtgt tgggggcagt ccccctaagt ccaagttact gtttagtaac acagcagctc 2760
aaaaattaag aggaatggat gaagtataca acctcttcta tgtcaacaac aactggtata 2820
tttttatgcg actgcaccag attctctgcc tgaggctgct acggatttgt tcccaagccg 2880
aacggcaaat tgaagaagaa aaccgagaga gagaatggga acgggaagtg ctgggcataa 2940
agcgagacaa gagtgacagc cctgccattc agctacgtct caaagaacct atggatgttg 3000
atgtagaaga ttattaccca gctttcctgg acatggtgcg gagcctgctg gatggcaaca 3060
tagactcatc acagtatgaa gattcactga gagagatgtt caccattcat gcctacattg 3120
cctttaccat ggacaaactg atccagagca ttgtcagaca gctgcagcat atcgtgagtg 3180
atgagatctg tgtgcaggtg actgaccttt acctggcaga aaataataat ggggccaccg 3240
gaggccagct gaacacacag aactcaagga gcctcctgga gtcaacgtat cagcggaaag 3300
ctgagcagct aatgtcagat gagaattgct ttaagcttat gtttattcag agccaaggcc 3360
aggtccagct gactattgag cttctggaca cagaagagga gaattcggat gaccctgtgg 3420
aagcagagcg ctggtcagac tacgtggagc gatacatgaa ttcagatact acctcgcctg 3480
agcttcgtga acatctagca cagaaaccag tatttctccc caggaatcta cggcggatcc 3540
ggaagtgtca acgtggtcga gagcagcagg aaaaggaagg gaaggaagga aacagcaaga 3600
agaccatgga gaatgtggat agtctggata agctggagtg tagattcaag ctgaattcct 3660
acaagatggt gtatgtgatc aaatcagagg actatatgta tcggaggacc gccctgctcc 3720
gggctcatca gtcccatgag cgtgtaagca agcgtctaca tcagagattc caggcctggg 3780
tagataaatg gaccaaggag catgtgcccc gtgaaatggc agcagagacc agcaagtggc 3840
tcatgggtga ggggctggag ggcctggtgc cctgtaccac cacctgtgat acagagaccc 3900
tgcatttgt gagcattaac aagtatcgtg tcaaatacgg cacagtattc aaagcccctt 3960
aactgcaaag ccagagcaga taacttgggg tgtgtgtggg gatgtgtgtg tgggcctatg 4020
cactcacaca ctgaagaaac aaggaagatg cctttcaagc ctcactgggc ctctctggga 4080
catggccacc tgacctgtgt gtggctggtg cagcctggca ccaagtgggc tacctgttag 4140
gaacatgaat accttacaaa gctgaaagct ggaactttcc ccaaagggtt tgggtatagc 4200
ctgcctggag gggacggaag tccatgcaag ccaaagacat gcaggttgct tgcacacaaa 4260
cagcaagagc taaaactgga ttctcctggg cctaactttc atgaagggac cggatgctgt 4320
tcacacttgg actgttccag cacagtggga gattcagatc tgcggctcca tagggagcgg 4380
ataatcgaag cagtactgat tactttccc 4409

<210> 44
<211> 1290
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4019390CB1

<400> 44
ccagcctctg tgtccctgtg acctgcagat attgggagat ccacagctaa gacgccggga 60
ctccctggaa acctagaaat ggaaccactg acatttaaag atgtcgccat agaattctct 120
ctggaggagt ggcaatgcct ggatactgca cagcgggatt tatataggaa tgtgttgtta 180
gagaactaca gaaacctggt ctttttgggt attgctgtct ctaagccata tctgatcacc 240
tgtctggagc aaaaaaaaga gccctggaat ataaaaagac atgagatggt agccaaaccc 300
ccagtaatgt cttttcattt tgcccaagac ctttggccag agcagaacat aaaagattct 360
ttccagaaag tgacactgag aagatacgga aaatgtgaat atgagaattt acagttaaga 420
aaaggctgta aacatgtgga tgagtgtacg gggcacaaag gaggtcataa tacagttaac 480
caatgtttga cagctacccc aagcaaaata ttccagtgta ataaatatgt gaaagtcttt 540
gataaatttt caaattcaaa tagatataag agaagacata caggaaacaa acacttcaaa 600
tgtaaagaat gtagcaaatc attttgcgtg ctttcacaac taactcagca tagaagaatt 660
catactagag tgaattccta caaatgtgaa gaatgtggaa aagcctttaa ctggttctca 720
actcttacta aacataagag aattcatact ggagaaaagc cctacaaatg tgaagaatgt 780
ggcaaagcct ttaaccagtc ctcacaactt actaggcata agataattca tactgaagag 840
aaacccaaca aatgtgaaga atgtggcaag gcctttaaac aggcctcaca ccttactata 900
cataaaataa ttcatactgg agaaaaacct tacaaatatg aagaatgtgg caaagtcttt 960
agccagtcct cacaccttac tacacaaaag atacttcaca ctggagagaa cctctacaag 1020

tgtaaagaat gtggaaaagc ttttaaccta ttctcaaatc ttactaacca taagagaatt 1080
catgctgggg agaaacccta caaatgtaaa gaatgtggca gagcttttaa catatcctca 1140
aaccttaata aacaggagaa aattcatact ggagggaaac tcaacaaatg tgaagaatgt 1200
gacaagcttt taaccgatcc ttaaaactta ctgcacataa gaaaattcta atggaagaga 1260
aaccctacaa atgtgaagaa tgtggcaagc 1290

<210> 45
<211> 1516
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 986452CB1

<400> 45
ggaccccaga gagccctgag cagccccacc gccgccgccg gcctagttac cgtcacaccc 60
cgggaggagc cgcagctgcc gcagccggcc gcagtcacca tcaccgcaac catgagcagc 120
gaggccgaga cccagcagcc gcccgccgcc cccgccctca gcgccgccga caccaagccc 180
ggcactacgg gcagcggcgc agggagcggt ggcccgggcg gcctcacatc ggcggcgcct 240
gccggcgggg acaagaaggt catcgcaacg aaggttttgg gaacagtaaa atggttcaat 300
gtaaggaacg gatatggttt catcaacagg aatgacacca aggaagatgt atttgtacac 360
cagggtgcgg aggcagcaaa tgttacaggt cctggtggtg ttccagttca aggcagtaaa 420
tatgcagcag accgtaacca ttatagacgc tatccacgtc gtaggggtcc tccacgcaat 480
taccagcaaa attaccagaa tagtgagagt ggggaaaaga acgagggatc ggagagtgct 540
cccgaaggcc aggcccaaca acgccggccc taccgcaggc gaaggttccc accttactac 600
atgcggagac cctatgggcg tcgaccacag tattccaacc ctcctgtgca gggagaagtg 660
atggagggtg ctgacaacca gggtgcagga gaacaaggta gaccagtgag gcagaatatg 720
tatcggggat atagaccacg attccgcagg ggccctcctc gccaaagaca gcctagagag 780
gacggcaatg aagaagataa agaaaatcaa ggagatgaga cccaaggtca gcagccacct 840
caacgtcggt accgccgcaa cttcaattac cgacgcagac gcccagaaaa ccctaaacca 900
caagatggca aagagacaaa agcagccgat ccaccagctg agaattcgtc cgctcccgag 960
gctgagcagg gcggggctga gtaaatgccg gcttaccatc tctaccatca tccggtttag 1020
tcatccaaca agaagaaata tgaaattcca gcaataagaa atgaacaaaa gattggagct 1080
gaagacctaa agtgcttgct ttttgcccgt tgaccagata aatagaacta tctgcattat 1140
ctatgcagca tggggttttt attatttta cctaaagacg tctcttttg gtaataacaa 1200
acgtgttttt taaaaaagcc tggtttttct caatacgcct ttaaaggttt ttaaattgtt 1260
tcatatctgg tcaagttgag atttttaaga acttcatttt taatttgtaa taaaagttta 1320
caacttgatt ttttcaaaaa agtcaacaaa ctgcaagcac ctgttaataa aggtcttaaa 1380
taatctcttg ttaggaaaaa tgtccattaa taaggccagt cttcagcaaa actaaaacca 1440
ttttgttcgt ttagctttcc tagtctgaca acgcaatact gttgaaccac agtcaaatat 1500
aatgacaaca ttggat 1516

<210> 46
<211> 5123
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2807579CB1

<400> 46
cggccccgga tgttggctgc tcttgctgcc gccgccgctc tcggggctgc tcggggcctc 60
gaccgctcgc agtaggcacc gttgggacct ttcccggggg ggcggagacg cacggagcgg 120
ccgggggctc ggccgggccg tcccggcctg cgcagcctga gattctcttt gtggccagat 180
gggtttgtat ggacaggctt gtccatctgt aacttcatta aggatgacat ctgaactgga 240
gagcagccta acgtctatgg actggttacc acagctcacc atgagagcag ccatccaaaa 300
atctgatgct acacaaaatg cacatggaac aggaatttct aagaagaatg cactccttga 360
cccaaataca actctggacc aggaagaagt ccaacagcac aaagatggga aacctccata 420
cagttatgcc agcctcatta catttgcaat taatagctca cccaaaaaga aaatgacttt 480
aagtgaaatt tatcagtgga tttgtgataa cttcccatat tatagagagg ctggcagtgg 540
ttggaagaat tccatacgac ataatctgtc attgaacaaa tgtttcctta aagtgcctcg 600

atctaaggat gaccctggaa aggggtccta ctgggcaata gacaccaatc cgaaggaaga 660
tgcgctgcct actcggccaa agaagagggc acgatctgta gaacgggtaa cattgtataa 720
cactgatcag gatggtagtg atagcccacg cagtagcctt aacaacagtc tctcagacca 780
gagtttggca tctgttaatt tgaacagtgt tggaagtgtg catagttata caccggtgac 840
aagccatcca gaatcagtct ctcaatcatt aactcctcag cagcaaccac agtacaacct 900
tccagaaaga gacaagcaac tacttttctc agaatataat tttgaagatc ttagtgcctc 960
atttcggagc ctttataagt cagtttttga gcagtcactt agtcaacaag gtttgatgaa 1020
catcccttct gaatcttccc agcagtccca cacttcatgt acctatcagc actctcccag 1080
cagtacagtg agcactcacc cacacagcaa ccaaagcagc ctgtccaaca gtcatggcag 1140
tggcctcaac accacaggca gtaattcggt tgcacaggtc tcactgtctc acccccagat 1200
gcacacacag ccatctccac atcctcccca tcgaccgcat ggtttaccgc agcatccgca 1260
gcgttcccca cacccagcac cacacccaca gcaacacagc cagctccagt cccctcaccc 1320
ccagcatccc tctccacatc aacacataca gcaccatccg aaccatcagc atcagacgtt 1380
aacacatcag gcacccccac ccccacaaca ggtatcctgt aattctggtg tttcaaatga 1440
ttggtatgcg acacttgata tgctaaaaga aagctgtcga attgccagca gtgttaattg 1500
gtcagatgta gacctttcac agtttcaagg tctgatggag agtatgagac aggcagatct 1560
caagaactgg tctttagacc aggttcagtt tgccgatctt tgttcttctc ttaatcagtt 1620
ctttacacaa actggcctta tacattcaca gagtaatgtt caacaaaatg tttgtcatgg 1680
tgccatgcat ccaacaaaac cttcccaaca cattggaaca ggaaatttgt acatagattc 1740
taggcaaaat ctccctcctt cagtgatgcc acccccctggt tatcctcata tcccacaggc 1800
actcagcact ccaggaacaa cgatggcagg ccatcacaga gccatgaacc agcagcacat 1860
gatgccttcc caagccttcc agatgcggcg ttccctgcct ccagatgaca tccaggatga 1920
ctttgattgg gattcaattg tgtagggctt gtttctgcaa gacaccagac cctaacgtta 1980
cctttctgtg cagtgaaggg aaaggtttaa gagaatccag ttgagaaaac aaacttgcta 2040
atcactttac caatgttatc aaaattactt ttgaagacaa tcagaaggat tttagctgga 2100
taacttactg cttttatctg acccaagcaa gtactacatg tttgtctccc tgccagctgc 2160
cctatgtagc tcctaactgt tgtgtgattt ggacggcttt ttgcatattt gtgtcagttt 2220
gatgttaacc acaagtgcca gactgatttt tcagacggag cctattttgc tgcaagcagt 2280
ttatataaag atacatatgt gtaaatatat gtacaaaaat tactgaaagg cttcagtttt 2340
ttctaattgg attattatgt cttgaaagtt attgtcagtt tttattcctt gttaggctat 2400
tttctgcagg atgcttttaa ctgatgtagg aaactgaaag gaaatagatt ttttccaaaa 2460
cccagttccc cttatttaat ctttttttaga aatgtgggta atgaattcta tctaataagt 2520
caaggaaacc agaatttgac acactccaac aatccaaagg ggcatgttgc tcctgagcag 2580
catgaagaac tgaccaaatt ggttttgatg cttgggggat catagagtat ttatgtctgc 2640
ttttctaaat ctgcattata atagctctaa aatttgttga ttggtaagaa attgggcatt 2700
gcttggctct ttaaacacat cagtgcttcc acattcacct atgtatttat tattcaaaag 2760
tgtcatttta atatttattg ctaccttctg tgaatgctta gctcctgtcg ggttcattaa 2820
ggagaaatgt gtctgaaagc acagaactat tattattttt ttctttttttt gagatggagt 2880
cttgccctgt tgcccaggct ggagtgcagt ggcacgatct cggctcactg caacctccgc 2940
ctcccgggtt caagcaattc tcctgcctca gcctcccgag tagctgggag tgcaggtgca 3000
caccgccatg cctggctgat tttttgtatt ttagtagaga tggggtttca ccatgttgcc 3060
tgggctggtt gcaaactctt gagctcaggc agtctgcctg cctcagcctc ctagagtgct 3120
gggattacag acatgagcca ctgcgcctgg cccagcgtta cttttcttga taagaattta 3180
cagataagac actcagagag acattgtgtc attctctatc atcactcctt ttaccatgtg 3240
aatagatttc tctgtcttgg cattcttgct gattcagact tattagttaa atcggatttt 3300
tctggaattt gaatcagatt tcatttgggc aacaacagca gggcagggtt tctaaagcag 3360
gtttccccac tcatctaggt tgtcttgaaa ggaggataga gccacttaca gttttatttg 3420
cttcagtgtt taatgtagat attatgtggc ctgctgtttc tgttgtcttg tatgtctgtg 3480
tatgcatgac atttggtccc tttctttgaa atgcagcccc tcttacgggt tttttagtgg 3540
tggtggtttg ttttgtttca tatcatatca cagtttgcat ggactgatta tcttagtttt 3600
atgataaaac tagaagaaat gagggttttt ttttaatgaa tagattttga attgattctt 3660
tagacccccc aaaaagtgtc agttctaatg gggaaatata ttccatcaag tcaaagagta 3720
ataactgctc actagttgct ttttagcatc tctggtctta tcagccatgc taaatcactt 3780
taattagcct tagtgattct atggttgagt aatctctact tgaactaaac aaacatcttt 3840
tgtttctgtg tgtgtgtgct gtgtgtgaga gtgtgcgcgc gcgtgtgtgt gtgtgtgttt 3900
taatggagtg ttgccttgaa tgaatcactg ggaagccagc catggtaagg gctggtgagg 3960
ttggggagaa aggaagagct ttatgtttct ctgttgtttg gaccctactt ggcatgaaaa 4020
aggaagctca gttccagccc cttggatcaa cgaaaatcag aggattctgg aaaggcagcc 4080
aacttgcgcc tcttagaagg atcagaggca agatgagatg gcagcctgca gagtaaatgc 4140
ttgaaaaaag aggggtgatt ttcaatggtt tgcttaagtc actgttttct agacaccaaa 4200
atagctgttt tgaaactgtt ttaattgctt gggtagcaat gtgcacttta aacaatttgg 4260
atattggaat gcagtctcat actgagtgat ttgagtagaa ccactgatga tgattttata 4320
aattgtgtga agcgaatttc ccatttggca atcatttact gatttgcagt gatgaatatt 4380

tttatgagaa tttaaactta gcaagaatgg ccatggaggc aaagccttca cccagaccca 4440
tcccactctc ctgtgatcca ggtggtccag gagcccagga caggcctttt ctgtgggccc 4500
tggccagaca gggttacctg gtgaggtgca gagagtccct ctagtggcca ttttgtatgg 4560
tagttgctaa tgcagaacaa gttctgtctt gggcttaaat tgactgaaga ctttaggggg 4620
aaagaatagt aaatgcatgt aaacaaatgg ggacactctg ttcaggagaa taatccgact 4680
ggcatttgtg gcagtttttg aaatgtaaat gtattcatgt gtgttcttgt aaatacgtgt 4740
cgctcagatg tcctttgaag tgggagggaa tcaatccggg gataatttca aatggaatag 4800
agtattttga tattgttcat tcagagggtg atgtgtacat atctatattg tatatatgtg 4860
atgaaaatgc attggctttt tgtgcagata caacctgctc tctgtactgc tgttggacag 4920
tcagtgtttt aatgtttcta cagttttgct attgcacgat ttcatatttt gcctctatga 4980
tgaacggcaa ccattatttg taactgttta gtgctgtaaa gaaatattcc aagtgtcatt 5040
aggattgttg ctgccagaac tgatatgcat gaatggcact taaaataaat atattatgtt 5100
aactctagaa aaaagaaaaa aaa 5123

<210> 47
<211> 707
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5724273CB1

<400> 47
ccaggctctt gtttcccatc agcctctggg tccaaactaa ccacaggtct ctgggtcctt 60
ttccggtgtc cttaaggttc ttggactctg cctgtgctgc caggttgcca gatttctaag 120
aaggagtttg aagaggagaa aaggattttt gcatgtttag aaatcaaggt tcaggagaat 180
cttgacagtg aatgtgtctc attacattga ataacagcta caggatgttc gagagctgtt 240
ctggcttttc tggattctgt gcttcaccaa tagaggaatc ccatggagca ttaattagtt 300
cttgcaattc cagaaccatg actgatgggt tggtgacatt cagggatgtg gccatcgact 360
tctctcagga ggagtgggaa tgcctggacc ctgctcagag ggacttgtac gtggatgtaa 420
tgttggagaa ctatagtaac ttggtgtcac tggatttgga gtcaaaaaca tatgagacca 480
aaaaaaatat tttcagaaaa tgatattttt gaaataaatt tttcccagtg gaaggacaaa 540
agtaaaaccc ttggccttga ggcatccatc ttcagaaata attggaagtg caaaagcata 600
ttcgagggac taaaaggaca tcaagaggga tacttcagtc aaatgataat cagctatgaa 660
aaaatacctt cttacagaaa aaagtaatct cttactccac atcaaag 707

<210> 48
<211> 2170
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3614884CB1

<400> 48
tgtttcaggt tcatgcatgt tgtagcatga gtcaatattt catttcttta aatttccaag 60
taagatttca ttgtatggat aggcgacaat ttgtccattt acccatagga aggcattttg 120
gttgtgtgca attttgggtt attatgaata atgcccctgt gaacattcaa gtgctaattc 180
acaacacatg ttcatactgt tgggtggcat cactgaccac cccttcatgt ctcttttttt 240
tttttgcctt ccatggcatg tttcaggcac aattctgtct tctcttgaac tgcataactg 300
agaactctgc aaattcccca aagaaggagg aaaaatgacc atgttcaagg aggcagtgac 360
cttcaaggat gtggctgtgg tcttcaccga ggaggagctg gggctgctgg acgtttccca 420
gaggaagctg tatcgagatg tcatgctgga gaacttcaga aacctgctat cagtggggca 480
tcaactttcc caccgagata cttttcactt ccaaagagaa gaaaagtttt ggatcatgga 540
gacagcaacc caaagagaag ggaattcagg aggcaagatc caaactgagt tggagtctgt 600
tccagaaaca ggaccacatg aagagtggtc ttgccagcaa atctgggaac aaactgcaag 660
tgagttaact agacctcaag actccataag tagctctcag ttctccacac aaggtgatgt 720
cccctcccag gttgacgcag gactatctat aattcacata ggagagacac cttctgagca 780
tgggaagtgt aaaaaattct tcagtgatgt ctccatcctt gatcttcatc aacaattaca 840
ctcaggaaag atatcccata catgtaatga gtacaggaag agattctgtt atagctcagc 900
tctttgtctt catcagaaag ttcacatggg agagaaacgc tataagtgtg atgtgtgtag 960

taaggcattt agtcagaact cacaactgca aactcatcag agaatccaca ctggagagaa 1020
accattcaaa tgtgagcagt gtgggaaaag tttcagccgt agatcaggaa tgtatgttca 1080
ttgcaaatta cacacaggag aaaaacctca tatttgtgag gaatgtggga aggccttcat 1140
tcacaattcc cagcttcggg aacatcaaag aatccatact ggggagaagc cattcaaatg 1200
ttatatatgt ggtaagagct tccatagtag atcaaatctt aataggcatt ccatggtcca 1260
catgcaagag aaatcattta gatgtgatac ctgtagtaat agctttggtc agagatcagc 1320
acttaatagt cattgcatgg accacacaaa agagaaacta tacaaatgtg aagaatgtgg 1380
aaggagcttc acttgtaggc aagatctttg taagcatcag atggaccata caggagacaa 1440
accatataat tgtaatgtat gtgggaaggg cttcaggtgg tcctcatgtc tttcaagaca 1500
tcagcgggtc cacaatggag aaacaacatt caagtgcgac ggatgtggga agagatttta 1560
tatgaattca cagggccatt cacatcagag agcctataga gaagaagaac tgtataaatg 1620
tcagaagtgt gggaagggct acattagtaa gtttaatctt gacttgcacc agagggtcca 1680
cacgggagag agaccttata attgtaagga atgtggaaag agcttcaggt gggcctcagg 1740
tattttgaga cataagagac tccatactgg agaaaaacca ttcaaatgtg aagagtgtgg 1800
gaagaggttt actgagaatt caaaacttcg tttccatcaa agaattcaca ctggagaaaa 1860
gccttacaaa tgtgaggagt gtggaaaggg attcagatgg gcctcaactc atctaaccca 1920
tcaaagactc cacagcagag aaaaactatt ccaatgtgag gattgtggga agagcagtga 1980
gcacagttca tgccttcaag accaacaaag cgaccacagt ggagaaaaaa catccaaatg 2040
tgaggactgt gggaagcgct acgagaggcg cttgaatcta gatatgattt tatcattatt 2100
tttaaatgat atataattat tgtccatatt tatggttaca gcatatttaa atacatgtat 2160
acaatgtatg 2170

<210> 49
<211> 2778
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3794954CB1

<400> 49
ggagagtcgc ttgaacctgg gaggtggggg ttgcagtggg ccaggattgt gctactgcac 60
tccagcctgg gcgacagtga gactccgtct caaaaaaaaa aaaatgctgt ggcaagggag 120
gcaagaaaga atgttttatc tgtagcattt tgtgctgata gaaaaggcag ggtccagtgg 180
aagaaggagg aaatggtaca taatttcttc ataaaaagag taaggccagg gagaagtgac 240
tgagtagaca cctgtagcta aacagaggtg gtgatgagtt tatggtgcct ccagcttctc 300
atggccccac ttgggaaatt cctccagtaa cattacactc actcactcct tgaccatcta 360
cctttactcc tctcctcatt tccctgttac caaaactatg tgccctaccc agggacttct 420
gggcagctct tttctcagat catttctcac cctctctcct ctactaactt agcctttcat 480
actgaatcac attttcctta aaaaacagct gctgacagaa gatgctacat tttgtttcag 540
ggcctggtaa ccatcaagga tgtgtcactg tgcttctctc aggaggagtg gcggagcctg 600
gacccctctc agacagactt ttatggagaa tatgtcatgc aggaaaactg tgggatagta 660
gtctctctga gtaagcatga ccttctccag ccactagaag attggactct gggaggacgg 720
agccatctgc ttcagggcta ttggatagtg tctgtctagg actctttatg tggcagggaa 780
tctgtctatg cttccgctgg agaaagaaca ctagggccaa tcccaagaga atgagagtaa 840
atatcattaa aaaaaaaaaa gtaaaaaagc aaatgatgct taattgtgga gtagaacttg 900
aggttcttct tggttgggaa ttaggaggtt ctcataaagc agagagaact ggaactgggt 960
ccttgtaaac ttgggaacca tgtagaggaa gatagggatg cagggagtag gactgtttta 1020
gtaggactaa agtgaagtta cgcccagaaa ggccagctgg gctggaagta tctgagtgga 1080
agaaaacagg ttcattggga aggagagtga ttttctaggc agagaagagg gacttaacag 1140
aaatgggtgg gggaagaggt tggggtgggg atatgtttct ttcttgacct tcttttgtct 1200
agagttccaa acaataggaa acagtttttt gtatttttt tcagaaatat agaattgctg 1260
ttgaactaat tgacagctct attttccttt cccatgagca ggatttccaa ttcccaaact 1320
ggacatgctt tctcaactag aaggaggaga agaacaatgg gtccctgacc cccaggactt 1380
agaggagagg gacattctga gggtcacata tacaggagat ggaagtgaac atgaggggga 1440
taccctgaa ctagaagcag aacctcccag aatgttatcc agcgtgtctg aagatactgt 1500
tctctggaac ccggagcatg atgagagctg ggattccatg cccagcagct ccagaggaat 1560
gctcctgggg ccccctttc ttcaggaaga tagtttctca aacctgctgt gtagcacaga 1620
gatggattcc ctgttaagac cccacacatg cccccagtgt gggaaacagt ttgtatgggg 1680
gtcccacctt gccaggcatc aacaaacaca cactggggag agaccctaca gctgcctcaa 1740
gtgtgagaag acctttgggc gaagacatca cctcatcagg caccagaaaa cccacctaca 1800
tgacaagacc agcaggtgct ctgagtgtgg taagaatttc cgatgcaact cccatctggc 1860

cagccaccag agagtgcatg cagaaggcaa atcctgcaaa ggccaagagg tggagagagc 1920
cctggcacaa ggaaacggcc gcgtgcccca ccagtgccaa agtgtcacgt gtgcactgaa 1980
tgtggggaag agctttggcc gaaggcacca ccttgtgaga cactggctga cccacactgg 2040
ggagaagccc ttccagtgcc ctcgctgtga gaagagcttt ggccgaaaac atcacctgga 2100
caggcacctg ctcacccacc agggacaaag tccccggaac agctgggaca gaggaacatc 2160
tgtcttttga aatctgtttc cactacagct atggtcaagt ctatcagccg gtgctaccag 2220
gagtcactgc cagggctgcc gttctcctga accccagtgg ccagaatcat aagccctgac 2280
cccatcccta gaaagatgag gtcccagcaa tggccagagc atttctcacc agttctgtga 2340
gatagcacat aaaaatagag ttctttgggc aaaacttttg ggaagcaatg catcctacat 2400
gggctgatat tcagcctgag ctgttctcta gaggagagtg gtactggcag tttatggctg 2460
aaatccattc tgattggttg gagtctatgc tataccagtt gttaaacatt ttgagtatca 2520
ctcttgcata ctgttactat tatattttct ctatatatag acagaaaggc cattttagaa 2580
tattaaaggc tctgaaaatt tctgcagtag acccaactga aggttctatt aaggcagtgt 2640
ttcctaaatg tatttgacct cagtagcctt tctatttaca tcccacagaa ctggtgacta 2700
tgaaacactc tggtgaccct gggttagaga gacccaggca acagagatag tgtccaggct 2760
cagagcccct tttaatcg 2778

<210> 50
<211> 2478
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7399016CB1

<400> 50
cgccggcgtc cccggcccgg cctccccggc cccgggctcg gccagaaggg acccggcgtg 60
aggagcgctg tcaccgcggg caggtcgccc tggggtgccc gcgcgtggga atcgccctcc 120
cctgcttcgt gccctcgcgg agccggaaag gaggcggcac ctgcaattgg cagacaggaa 180
accgcggcgt ttcctggagg gcgcagcgcc tgccccgggt ctcgcggtcg ctgggaggag 240
ggtgcagttc caggtttcac tcttgaagtt gcttcgctgg agaggcctta gaggcctttg 300
aaatttgaaa gatttcttgg accagcggcc acatccccca cagggcggct tgtcccgccg 360
agagccgcgc gcccgagccg cttgctgtgt ccgggagccg cgcgaagggc gctgcaggcg 420
ctgccctcgg acctggagtc cgggacgccc tgtgctcagg agcctccttt gccagctgct 480
aacacttcct gccgctctgt gcaggagcag gccgggctct cgccatgggt cactgtcgcc 540
tctgccacgg gaagttttcc tcgagaagcc tgcgcagcat ctccgagagg gcgcctggag 600
cgagcatgga gaggccatcc gcagaggagc gcgtgctcgt acgggacttc cagcgcctgc 660
ttggtgtggc tgtccgccag gaccccacct tgtctccgtt tgtctgcaag agctgccacg 720
cccagttcta ccagtgccac agccttctca agtccttcct gcagagggtc aacgcctccc 780
cggctggtcg ccggaagcct tgtgcaaagg tcggtgccca gcccccaaca ggggcagagg 840
agggagcgtg tctggtggat ctgatcacat ccagccccca gtgcctgcac ggcttggtgg 900
ggtgggtgca tggacatgcg gccagctgcg gggccctgcc ccaccttcag aggacactgt 960
cctccgagta ctgcggcgtc atccaggtcg tgtggggctg cgaccagggc cacgactaca 1020
ccatggatac cagctccagc tgcaaggcct tcttgctgga cagtgcgctg gcagtcaagt 1080
ggccatggga caaagagacg gcgccacggc tgccccagca ccgagggtgg aaccctgggg 1140
atgcccctca gacctcccag ggtagaggga cagggacccc agttggggct gagaccaaga 1200
ccctgcccag cacggatgtg gcccagcctc cttcggacag cgacgcggtg gggcccaggt 1260
cgggcttccc acctcagcca agcctgcccc tttgcagggc cccagggcag ttgggtgaga 1320
agcagcttcc atcttcaacc tcggatgatc gggtaaaaga cgagttcagt gacctttctg 1380
agggagacgt cttgagtgaa gatgaaaatg acaagaagca aaatgcccag tcttcggacg 1440
agtcctttga gccttaccca gaaaggaaag tctctggtaa gaagagtgaa agcaaagaag 1500
ccaagaagtc tgaagaacca agaattcgga agaagccggg acccaagccc ggatggaaga 1560
agaagcttcg ttgtgagagg gaggagcttc ccaccatcta caagtgtcct taccagggct 1620
gcacggccgt gtaccgaggc gctgacggca tgaagaagca catcaaggag caccacgagg 1680
aggtccggga gcggccctgc ccccaccctg gctgcaacaa ggttttcatg atcgaccgct 1740
acctgcagcg ccacgtgaag ctcatccaca cagaggtgcg gaactatatc tgtgacgaat 1800
gtggacaaac cttcaagcag cggaagcacc ttctcgtcca ccaaatgcga cattcgggag 1860
ccaagccttt gcagtgtgag gtctgtgggt tccagtgcag gcagcgggca tccctcaagt 1920
accacatgac caaacacaag gctgagactg agctggactt tgcctgtgac cagtgtggcc 1980
ggcggtttga gaaggcccac aacctcaatg tacacatgtc catggtgcac ccgctgacac 2040
agacccagga caaggccctg cccctggagg cggaaccacc acctgggcca ccgagccccc 2100
ctgtgaccac agagggccag gcggtgaagc ccgaacccac ctgaggacgg cagtgaggat 2160

gagcacctct agcagcctgg actccgcagt ggctgtgtca gcctcaccct tcgtgtgcac 2220
ccgcatggga gggtcggagg gtgctgcccg cccttggtgc tggaggcggg cttggtgtcc 2280
ggctcaagta gccttcctct gctctgggac cagtggttta ttttcccgca aacgctgagt 2340
gactcggggc cggacagttc ataaataatt gattcctttc cccactaaag cagtcgagga 2400
gatttgtaat ccacttttta gtgcaacaag agctccatgt tatgcttgta ataaattatt 2460
tacacgggaa aaaaaaaa 2478

<210> 51
<211> 1947
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6996690CB1

<400> 51
gtggtggagt ccaccctggc tgtggaccta ttcaaactca gtctgtctct tgactggagg 60
aggagaagga aaagttgagc acattgccaa ccggagttcc ccaagaatgg caagttaaaa 120
ccaaagtgtc agcacctcaa caacattttt ttttggctta acaaaacatc aaacggcata 180
caaatggcag aaatccacaa tggaggggaa ctctgtgact ttatggaaaa tggagaaatc 240
ttcagtgaac actcatgcct taatgcacac atgggaactg aaaatacagg ggacacttat 300
gactgtgatg agtatggaga aaactttccc atgttacaca acagtgcccc tgctggagag 360
acactttctg tgttgaatca gtgcagaaaa gccttcagcc tgccaccaaa tgttcaccag 420
agaacgtgga taggagacaa atcctttgaa tacagtgact gtgaggaagc ctttgttgat 480
cagtcacatc ttcaggcaaa taggataact cacaatggag aaacactcta tgaacagaag 540
caatgtggga gagcttttac ttactccaca agccatgctg tgtctgttaa aatgcatact 600
gtagaaaaac cctacgaatg taaggaatgt ggaaaattct ttagatattc ttcatatctt 660
aatagtcaca tgcgaaccca tactggggag aaaccctatg aatgtaagga atgtgggaaa 720
tgcttcactg tttcttcaca cctagttgaa catgtaagaa ttcatactgg agagaaacct 780
tatcaatgta aggaatgtgg aagagccttc gctgggcgct caggccttac taaacatgta 840
cgaatacaca ctggagagaa gccctatgaa tgtaacgaat gtgggaaagc ctacaatagg 900
ttttatctac taactgaaca ttttaaaact cacacagagg agaagccctt tgaatgtaag 960
gtatgtggaa aatccttcag aagctcttca tgccttaaga atcactttag aattcacact 1020
ggaataaaac cctataaatg taaggagtgt gggaaagcat tcactgtttc ctcaagctta 1080
cataatcatg taaaaatcca tactggagag aagccctatg aatgtaagga ctgtgggaaa 1140
gcctttgcta catcttcaca actcattgaa catataagaa ctcacactgg agagaaaccg 1200
tatatatgta aggaatgtgg gaaaaccttc cgtgcttctt cacatctaca gaaacatgtt 1260
agaattcaca ctggagagaa accctatata tgtaacgaat gtgggaaagc ctacaataga 1320
ttttatttac ttactaaaca tttaaaaaca cactgaagaa aaactctgaa tgtaaggaat 1380
gtggaaatct tttaggaatt tctcatacca taattaccac atttgaattc acactgaaga 1440
gaaattttat tggtacaaag aatgtgagga ggccttcagt tatccccatt cacttcgaag 1500
acatgaaaga actcatactc tatgtgtccc ctataaattt taggaatgtt agggagccaa 1560
tggaatcttt ttttattatt ttatttttatt ttatttttatt ttatttttatt ttatttttatt 1620
ttatttttatt ttatttgaga cagagttttg ctgttgtcat cacacaggct ggagtgcaat 1680
ggcatgatcc cagctcactg caacctttgt ctcctgggtt caggagaatt gcttgaaccc 1740
aggaggtgga ggttgcagtg agccaagatc gcaccaccgc atgagagaga gagattacta 1800
tttcttgtcc ctttttctca gtttgattat atttatatac atatgtcagt aaatctgttt 1860
tcagtattga tgtttaataa agaatgtaca atggcaaaaa aaaaaaaaaa aaaaaaaaaa 1920
aaaaaaaaaa aaaagaaaga aaaagaa 1947

<210> 52
<211> 3553
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7740866CB1

<400> 52
ggcgcggggt ccaagatggt ggcgctagga gccgcgaccc agtgatagcg gccgtggagg 60
ggcccccgac cgagcgggag gttgggggta gcctggaggt gagaccccgc tgcgttcaca 120

gagctggcgg ccgcggcgcc tgcgtccacc ggcaagcgag gagtggagca gagctcatat 180
gcctatgggg agaggcctgg gggaccgcag gaggatgtag gcgccgggtg cgcatgactg 240
ggccttcttg ctctcggtcg gcttcttggt ctcggcgtgc cccattcccg gccctgtccc 300
gctactcttg cctgttccag ttccctcctg tgggtaccgc taacggcatc ttcccgaggc 360
ccagaatcca ccgtcaggca taaattgatt attcctcgcc cacatctgtg ccagattctg 420
aagacaggaa taagatgaag gaatggaaat caaagatgga aatttctgaa gaaaagaagt 480
cagcaagggc tgcatccgaa aaactccaaa gacaaatcac ccaggaatgt gagttagttg 540
aaaccagtaa ttctgaggac agattattga agcactgggt aagcccttta aaggatgcaa 600
tgagacatct cccttcccaa gagagcggta tcagggaaat gcatattatc ccccagaaag 660
ccattgtggg agagattggc catggatgta atgaaggaga aaaaatactt tctgcaggag 720
aaagctccca tagatatgag gttagtggcc aaaacttcaa acagaagtca ggattaactg 780
aacatcagaa aattcataat ataaataaga cctatgaatg taaggaatgt gaaaaaacct 840
tcaacaggag ttcaaacctg atcatacatc agagaattca tacaggaaat aagccatatg 900
tgtgtaatga atgtgggaaa gactctaatc aaagttcaaa tcttattata catcagagaa 960
ttcatacagg aaagaaacct tatatatgtc atgaatgtgg aaaagacttc aatcagagct 1020
ccaatctggt gagacataag caaattcaca gtggtgggaa tccctatgag tgcaaagagt 1080
gtgggaaggc ttttaaggga agctcaaacc ttgtcctgca ccagagaatc cacagtaggg 1140
ggaagccata tttatgcaat aaatgtggga aggctttcag tcaaagcaca gatcttatta 1200
tacatcacag aattcacact ggagagaaac cctatgaatg ttatgactgt ggacagatgt 1260
tcagtcaaag ttcacacctt gtcccacatc agagaattca cactggagag aaacccctca 1320
aatgtaatga atgtgaaaaa gccttcaggc agcattctca ccttactgaa caccagagac 1380
tccacagtgg agagaaaccc tatgaatgtc acagatgtgg gaagaccttc agtgggcgca 1440
cagcttttct taaacatcag agattgcatg ctggagagaa aattgaagaa tgtgagaaaa 1500
ccttcagcaa ggatgaggag cttagggaag agcagagaat tcaccaggaa gagaaagctt 1560
attggtgtaa tcagtgtggt aggaatttcc agggcacctc agacctcatc agacatcagg 1620
taactcatac aggagagaaa ccatatgaat gtaaagaatg tgggaaaact ttcaatcaga 1680
gctcagacct tctgagacat catagaattc acagtggaga aaaaccttgt gtatgtagca 1740
aatgtgggaa atcttttagg ggcagctcag atcttattag acaccatcgt gttcatactg 1800
gagagaaacc ctatgaatgt agtgaatgtg ggaaagcctt tagccagagg tcacaccttg 1860
ttacacacca gaaaatccat actggagaga agccctatca gtgcactgaa tgtgggaaag 1920
ccttcaggcg gcgttcactc cttattcaac atcggagaat tcatagtggt gagaaaccct 1980
atgaatgtaa ggaatgtggg aagctcttca tttggcgcac agctttcctc aaacatcaga 2040
gcctgcatac tggagagaaa cttgaatgtg agaaaacctt cagccaggat gaagagctta 2100
ggggagagca gaaaattcac caggaagcga aagcttattg gtgtaatcag tgtggtaggg 2160
ctttccaggg cagctcagac ctcatcagac atcaggtaac tcatacaaga gagaaaccat 2220
atgaatgcaa agaatgtggg aaaactttca atcagagctc agaccttctg agacatcata 2280
gaattcacag tggagaaaaa ccttatgtat gcaacaaatg tgggaaatct tttaggggta 2340
gctcagatct tattaaacac catcgtattc atactggaga gaaaccctat gaatgtagtg 2400
aatgtgggaa agccttcagc cagaggtcac accttgctac acaccagaaa atccatactg 2460
gagagaaacc ctatcagtgc agtgaatgtg ggaatgcctt caggcggcgt tccctcctta 2520
ttcaacatcg gagacttcat agtggtgaga aaccctatga atgtaaggaa tgtgggaaac 2580
tcttcatgtg gcacacggct ttcctcaaac atcagagact gcatgctgga gagaaacttg 2640
aagaatgtga gaaaaccttc agcaaggatg aggagcttag aaaagagcag agaactcacc 2700
aggaaaagaa agtttattgg tgtaatcagt gtagtaggac cttccagggc agctcagatc 2760
tcatcagaca tcaggtaact catacaagag agaaaccata tgaatgtaaa gaatgtggga 2820
aaactcaatc agagctcaga ccttctgaga catcatagaa ttcacagtgg agaaaaacct 2880
tacgtatgca ataaatgtgg ggaatctttt aggagcagct cagatcttat taaacaccat 2940
cgtgttcata ctggagagaa acctcatgaa tgtagtgaat gtgggaaagt ctttagccag 3000
aggtcccacc ttgtcacaca ccagaaaatc cacactggag agaagcccta tcagtgcact 3060
gaatgtgaaa aagccttcag gcggcgttca ctccttattc aacgtcggag aattcatagt 3120
ggtgagaaac cctatgaatg taaggaatgt gggaaactct tcatgtggca cacagctttc 3180
ctcaaacatc agagactgca tgctggagag aaacttgaag aatgtgagaa aaccttcagc 3240
aaggatgagg agcttagggg agagcagaaa attcaccaag aagagaaagc ttattggtgt 3300
aatcagtgtg gtagggcttc caggggggcaa ggctcagacc tcatcggaca tcaggtaact 3360
cataccagga gagaaaacca tatgaatgtt aacgaatgtg gggaaaactt ttcaatcaga 3420
ggtcagactt cttgagacac tcctcgaatc caggggggaaa aactttgttg gaccatgtgg 3480
gaccctaggg cgccaatttt tacccccgtc tctgggaacc ttaaccggtg gaggctaaaa 3540
acggtgaaat aaa 3553

<210> 53
<211> 1760
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 8181605CB1

<400> 53
actctcccag attgaacaag ggaaggagca ctgcaactgg cgccgcctgg ccccaagatt 60
ccagatgttc ctgatggacc ccagtccagg ctcggggccc ccagttcccg ccccagacct 120
cttgatgcag atcaagcagg agggtgagct ccagctccag gagcagcagg ccctgggcgt 180
ggaggcgtgg gcagccgggc agccagatat tggggaggag ccctggggcc tcagccagct 240
ggattccgga gcaggagaca tctccacgga tgccacctct ggtgtccatt ccaacttttc 300
caccaccatc ccgcccacct cctggcaaac ggatctccct ccccaccatc cctcttcagc 360
atgctcggac gggaccctga agctcaacac agcagcctcc acggaagcag atgtaaaaat 420
tgtaataaaa acagaagtcc aggaagagga ggtggtggcc acacccgtac atcctactga 480
cctagaggct cacgggaccc tgtttggacc aggccaagcc acacggttct tccctagtcc 540
tgcccaggaa ggagcctggg aaagccaggg caagctcctt ccccagccag gaccctgtgc 600
tggggctgcg agagcccgcc cggcctgaga gggacatggg tgagctcagt cctgctgtgg 660
cccaggagga gacccctcct ggggactggc tcttcggagg ggtccggtgg ggctggaatt 720
tccggtgtaa accgccagtg ggcctgaacc cgaggacggg gcccgagggg cttccttact 780
cctccccgga caacggagag gccatcttgg accccagcca ggccccaagg ccattcaacg 840
aaccctgtaa ataccctggc cggaccaaag gctttggcca caagccaggg ctgaagaagc 900
accccgcggc gccccccggg ggccggccct tcacctgcgc cacgtgtggg aagagcttcc 960
agctgcaagt cagcctgagc gcgcaccagc gcagctgtgg ggcgcccgac gggtcgggcc 1020
cgggcacagg cggtggcggc agcggcagtg gcggcggcgg tggcggcagc ggtgggggca 1080
gcgcacggga tggcagcgcc cttcggtgtg gggagtgcgg ccgttgcttc acgcgccccg 1140
cgcacctcat ccgccatcgc atgctgcaca ccggcgagcg gcccttcccc tgcaccgagt 1200
gtgagaagcg cttcaccgaa cgctccaagc tcatcgacca ctaccgaacg cacacgggcg 1260
tgcggccctt cacctgcacc gtctgcggca aaagcttcat ccgcaaggac cacctccgca 1320
agcaccagcg caaccatgca gcgggcgcca agaccccggc ccgaggccag ccactcccga 1380
cgccgcccgc acctcctgat cccttcaaga gccccgcctc caaaggacct ttggcctcca 1440
cagaccttgt gaccgactgg acttgtggcc tcagcgtcct gggacccacc gatggcgggg 1500
acatgtgagc gcctccagcc ccatagcccc tgccggccgc acgtgtaaaa agccccgtgt 1560
gcaggcagca gggcggcgtg gaagcttcag gcagacgcgg gacggggaga accaaatgtc 1620
caagttgctg aactgatgat cactggagaa agagaaacta tccaccagga cagctgccac 1680
ctctaaacca agtagaattc tctgtgaaat gggagcgccg tagaatttta agtaatttaa 1740
ttgcaacctg ttttgttttt 1760

<210> 54
<211> 2772
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 8266487CB1

<400> 54
gtagcgaatt atgttatact ccccttagac ccatgcacca tttaccaact aatgctagaa 60
acttaaagct tgtttccagc caaatacctc ttcttttatc acaaatgccc tagtgtatag 120
ctgttgagag agcagcactg agcactctgg gtagcctgtg gctttttatt tccagagtgt 180
ggagatttcc atctcaagtg tgaagacagc tcatggtttc aggaaaagcc cagagtttca 240
gacctgctgt tctctcaagt ctcctggtat aatgtgtcaa tcagttccca acctatggac 300
tgtgagctca tagggttcac agtggtcact ccagcgctct gtgaatgtgc ccaactgtgc 360
cctgcctgtg cactaaatga tgaatatatc acacaataat aaaaacaact accatgtgac 420
catctaccat ttacgcagac acaatcctag ggaagaaaga cggtgggtat atagtgaata 480
tcggttgaag ggaattaata tctgatcttc tcaattaacg ggtattaaaa gtacaactat 540
tatcatacac ccattcttga tatttttga aattagaaaa ggtttcttca acctgaaaag 600
acggctccgg ccccagcacg cccggctacc gctgccccga gccgcagtgc gcgctggcct 660
tcgccaagaa gcaccagctc aaggtgcacc tgctcacgca cggcggcggt cagggccggc 720
ggcccttcaa gtgcccactg gagggctgtg gttgggcctt cacaacgtcc tacaagctca 780
agcggcacct gcagtcgcac gacaagctgc ggcccttcgg ctgtccagtg ggcggctgtg 840
gcaagaagtt cactacggtc tataacctca aggcgcacat gaagggccac gagcaggaga 900
gcctgttcaa gtgcgaggtg tgcgccgagc gcttccccac gcacgccaag ctcagctccc 960
accagcgcag ccacttcgag cccgagcgcc cttacaagtg tgactttccc ggctgtgaga 1020

agacatttat cacagtgagt gccctgtttt cccataaccg agcccacttc agggaacaag 1080
agctcttttc ctgctccttt cctgggtgca gcaagcagta tgataaagcc tgtcggctga 1140
aaattcacct gcggagccat acaggtgaaa gaccatttat ttgtgactct gacagctgtg 1200
gctggacctt caccagcatg tccaaacttc taaggcacag aaggaaacat gacgatgacc 1260
ggaggtttac ctgccctgtc gagggctgtg ggaaatcatt caccagagca gagcatctga 1320
aaggccacag cataacccac ctaggcacaa agccgttcga gtgtcctgtg gaaggatgtt 1380
gcgcgaggtt ctccgctcgt agcagtctgt acattcactc taagaaacac gtgcaggatg 1440
tgggtgctcc gaaaagccgt tgcccagttt ctacctgcaa cagactcttc acctccaagc 1500
acagcatgaa ggcgcacatg gtcagacagc acagccggcg ccaagatctc ttacctcagc 1560
tagaagctcc gagttctctt actcccagca gtgaactcag cagcccaggc caaagtgagc 1620
tcactaacat ggatcttgct gcactcttct ctgacacacc tgccaatgct agtggttctg 1680
caggtgggtc ggatgaggct ctgaactccg gaatcctgac tattgacgtc acttctgtga 1740
gctcctctct gggagggaac ctccctgcta ataatagctc cctagggccg atggaacccc 1800
tggtcctggt ggcccacagt gatattcccc caagcctgga cagccctctg gttctcggga 1860
cagcagccac ggttctgcag cagggcagct tcagtgtgga tgacgtgcag actgtgagtg 1920
caggagcatt aggctgtctg gtggctctgc ccatgaagaa cttgagtgac gacccactgg 1980
ctttgacctc caatagtaac ttagcagcac atatcaccac accgacctct tcgagcaccc 2040
cccgagaaaa tgccagtgtc ccggaactgc tggctccaat caaggtggag ccggactcgc 2100
cttctcgccc aggagcagtt gggcagcagg aaggaagcca tgggctgccc cagtccacgt 2160
tgcccagtcc agcagagcag cacggtgccc aggacacaga gctcagtgca ggcactggca 2220
acttctattt ggtatgaagc actctattca gtcaccacca tataggtcac ttctctcata 2280
ctcggtcttg aggatattct ggattaatcc tttctatgca gacgtttctg gtttacaaaa 2340
ggacgcagcc ctggactaca agtctggaac tgacaagttc ttatgacctt gacaaatcac 2400
cttaacccat ctgagcctta aattctcatt tatttcctgc ataaggagat ttggctaaat 2460
gctttctgag gtcctttgga gtcctgtagc tccatggtaa tgtgctcctt tccttgaaga 2520
ctgggggttt tgtaatgttg agatactttg cctctatgct tctcagctca tgaccagtcc 2580
tagaagagga gtcgagacat aagccacctt cagaggttca atggaaactt taaaaccata 2640
ccaaactctt ttttaaaatt agaattaaca aaaaaaaaaa aaagggtggg gtttatgagc 2700
cttagttctt ggaggattat aagagtactt ccccagtttt gaggctggac agttaatata 2760
ctttatatca at 2772

<210> 55
<211> 1151
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5552784CB1

<400> 55
gaggtcgcgt agggcctatt atgatgattt ctacaggagg ttgaagagat aagacccttc 60
cctgtgctcc ccccccccca ctccttaatt acggattgag caggggaggg gccggtgggg 120
ctcaggtgag cacacaggga gaaagggacg tgggcggggc cttacagagg gtgagcgaat 180
ccgaaaagac ctagaacctc gttgctggga gacaagtccc gccctgcagg cggcaccgga 240
agtggccggc tgggatcagc ctttaagatg gcgtctcctc agggggggca gattgcgatc 300
gcgatgaggc ttcggaacca gctccagtca gtgtacaaga tggacccgct acggaacgag 360
gaggaggttc gagtgaagat caaagacttg aatgaacaca ttgtttgctg cctatgcgcc 420
ggctacttcg tggatgccac caccatcaca gagtgtcttc atactttctg caagagttgt 480
attgtgaagt acctccaaac tagcaagtac tgccccatgt gcaacattaa gatccacgag 540
acacagccac tgctcaacct caaactggac cgggtcatgc aggacatcgt gtataagctg 600
gtgcctggct tgcaagacag tgaagagaaa cggattcggg aattctacca gtcccgaggt 660
ttggaccggg tcacccagcc cactggggaa gagccagcac tgagcaacct cggcctcccc 720
ttcagcagct ttgaccactc taaagcccac tactatcgct atgatgagca gttgaacctg 780
tgcctggagc ggctgagttc tggcaaagac aagaataaaa gcgtcctgca gaacaagtat 840
gtccgatgtt ctgttagagc tgaggtacgc catctccgga gggtcctgtg tcaccgcttg 900
atgctaaacc ctcagcatgt gcagctcctt tttgacaatg aagttctccc tgatcacatg 960
acaatgaagc agatatggct ctcccgctgg ttcggcaagc catccccttt gcttttacaa 1020
tacagtgtga aagagaagag gaggtagggg ccaagccccc accccatccc actccccttc 1080
cctccccaga tatttatgtg aaatgaactg cagctttatt ttttgaaata aaaactttta 1140
aaaaccaaaa a 1151

<210> 56

<211> 2230
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7281230CB1

<400> 56
ggcacctcag tgtgtccctc aggcacctgg acacacaggt gggcaaagat gctgacactg 60
ggggccagaa acgccgatac ccccattctt ccatttttac tcagtctcag gtggtcccaa 120
ggaggggacc ttcgtggcct tgctgtgctc acaggctcac ctttgtgttt ccagctgcag 180
agaaagtacc ctgggccatg cagctgcact cccctcccag gaaaggggca ggatggctgc 240
ccagatgagt gaggcatcag ccctggcccc ccaggtcttc ccgagtccac tggaactgat 300
ggtgggtgag cccagctcca agagccctgg ccagtgcttc tgggggttct gctatgagaa 360
ggcagcaggg ccccgagggg ccctggctca gctgcgtgag ctgtgttgcc agtggctgat 420
gcctgaggcc tgctccaagg agcagatgct ggagctcttg gtgctggagc agttattggg 480
cacactgctc ccagagatcc aggcctacac gcaggagcag tggctaggca gccctgagga 540
ggccactgcc ctggcagagc ggctacagca ggagtcagct gggccaggac tccagatgag 600
tgggggctgg tctgggggct gggtgccagc ccccaggccc caagaggagc tggtccccag 660
gacagaggag ggagaggagc aagaggctcc cctgggcccc ttccaggccc cacctccagg 720
tcacaggcgc gagatggagt ccccaagagg gtggaccctg caggtggccc cagaggaagg 780
ccaggtcctc tgcaatgtga agactgccac gaggggcctc tctgaggggg ctgtgtctgg 840
aggctggggg gcctgggaaa actccacgga ggttccgagg gaggcagggg acggccagcg 900
gcagcaagcc acactggggg cggcggacga acagggaggc cccggcaggg agctgggccc 960
ccgcagacgg tgggcgggac ggggctgggc ccaggagcga gcctgcagac cgggcgttgc 1020
gcccttcgcc tctccccaga ggagccgggc tgccggtgcg gggagtgcgg caaggcgttc 1080
agccagggct cttacttgct gcagctcggc gcgtgcccca ggcgagaagc cgtacacgtg 1140
ccccgagtgc ggcaaggcct tcgcctggag ctccaacctc agccagcacc agcgcatcca 1200
cagcggcgag aagccctacg cttgcaggga gtgcggcaag gccttccgcg cgcactcgca 1260
gctcatccac caccaggaga cacacagcgg cctgaagccc ttccgctgcc cggactgcgg 1320
caagtccttc ggccgaagca ccacgctggt gcagcaccga cgcacgcaca cgggcgagaa 1380
gccctacgag tgcccggagt gcggcaaggc cttcagctgg aactccaatt tcctggagca 1440
ccggcgcgtg cacacgggcg cgcggccgca cgcctgccgg gactgtggca aggccttcag 1500
ccagagctcc aacctggccg agcacctgaa gatccacgcg ggcgcacggc cacacgcctg 1560
tcccgactgc ggcaaggcct tcgtgcgtgt ggcggggctg cggcagcacc ggcgcacgca 1620
cagcagcgag aagcccttcc cctgcgccga gtgcggaaag gctttccgcg agagctcgca 1680
gctcctgcag caccagcgca cgcacactgg tgagcggccc ttcgagtgcg ccgagtgcgg 1740
ccaggctttc gtcatgggct cctacctggc ggagcaccgg cgcgtgcaca cgggcgagaa 1800
gcctcatgcg tgcgcccagt gcggcaaggc cttcagccag cgctccaacc tactgagcca 1860
ccggcgcacg cactcgggcg ccaagccctt cgcctgcgcc gactgcggca aggccttccg 1920
cggcagttcc ggcctggcgc accaccggct ttcgcacacg ggagagcgac ccttcgcctg 1980
cgcagaatgc ggcaaggcct tccgcggcag ctccgagctg cgccagcacc agcgcctgca 2040
ctctggcgag aggccgttcg tctgcgccca ctgcagcaag gccttcgtgc gcaagtcgga 2100
gctcttaagc caccggcgca cgcacacggg cgagaggccc tacgcttgcg gcgagtgcgg 2160
gaagcctttc agccaccgtt gcaacctcaa cgagcaccag aagcggcacg ggggccgcgc 2220
tgcgccctga 2230

<210> 57
<211> 1976
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7488424CB1

<400> 57
ggagcgcact tagaggaagc tgtgttttgg tgacctctga aactcagtac tgcggcgaat 60
gagctcctga ccttgaggag tacttaacag aattatttct cgaagaatca ttgtgggaac 120
ccttcaaaga acccagcgaa acatgaattc tggaatctcg caagtcttcc agagggaact 180
cacctgcccc atctgcctga actacttcat agacccggtc accatagact gtgggcacag 240
cttttgcagg ccctgtttct acctcaactg gcaagacatc ccaattctta ctcagtgctt 300

tgaatgccta aagacaacac agcagagaaa cctcaaaact aacattcgat tgaagaagat 360
ggcttcccgt gccagaaaag ccagtctctg gctattcctg agctctgagg agcaaatgtg 420
tggcactcac agggagacaa agaagatatt ctgtgaagtg gacaggagcc tgctctgttt 480
gctgtgctcc agctctctgg agcaccggta tcacagacac tgtcccgctg agtgggctgc 540
tgaggaacac cgggagaagc ttttaaagaa aatgcagtct ttatgggaaa aagtttgtga 600
aaatcagaga aacctgaacg tggaaaccac cagaatcagc cactggaagg attatgtgaa 660
tgtaaggcta gaagctatta gagctgagta tcagaagatg cctgcatttc atcatgaaga 720
agaaaaacat aatttggaga tgctgaaaaa gaaggggaaa gaaatttttc atcgacttca 780
tttaagtaaa gccaaaatgg ctcacaggag ggagatttta agaggaacgt atgcggagct 840
gatgaaaatg tgccataaac cagatgtgga gctacttcag gcttttggag acatattaca 900
caggagtgag tccgtgctgc tgcacatgcc ccagcctctg aatctagagc tcagggcagg 960
gcccatcact ggactgaggg acaggctcaa ccaattccga gtggatatta ctctgcctca 1020
taatgaagcc aacagtcata tcttccgacg tggagatttg agaagcattt gtattggatg 1080
tgaccgtcaa aatgcgcccc atatcactgc aacacctaca agttttcttg catggggtgc 1140
tcagactttc acctctggca aatattactg ggaggtccat gtgggggact cttggaattg 1200
ggcctttggt gtctgtaata agtattggaa agggacgaat cagaatggca atatacatgg 1260
agaggaggga ctctttagtc ttgggtgtgt taagaatgac attcagtgca atctctttac 1320
cacctcccca gttacactgc agtatgtccc aagacctacc aaccatgtag gattattcct 1380
ggattgtgaa gctagaactg tgagcttcgt tgatgttaat caaagctccc ctatatacac 1440
catccctaat tgctccttct cacctcctct caggcctatc ttttgctgta ttcatctctg 1500
accagagaca aatcagaaat gtgtttatct gctgtgggaa ccccttttatc ccataaagcc 1560
ctcttcctta tgccttatca aacaggacaa ataggttctg ttttatgtct tgaattgcat 1620
tctaatgtta ttaaaactca tttattgtgt tactattaaa tatgctgaaa acgctaaaag 1680
tatacgtatt ggttctttat taaataattt ttgaaaaatc attattcatg atcatggcat 1740
acagtatatt ctctttttt tctttattta tgactgtcac tgagtgaaat aatagatgac 1800
agacatgtct gaatgaagta aaaatcaatg gaagacagtc gggatctttt gcttcatgca 1860
aaaaacttgg agtgaagtct caatgataac tgggaaatgt ttttcttcct ctttatctaa 1920
ctatattaca cttatccatc aggtttcatt gtattaatct atcctttgag gtaata 1976

<210> 58
<211> 1357
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7487110CB1

<400> 58
atgacaatgg aaggggccag cgggtcgagt tttggaatag acacgatttt gtccagtgcc 60
agttcaggca gcccaggcat gatgaatgga gatttccgcc cgctcggtga ggccaggacc 120
gcggatttta ggagtcaggc caccccatct ccctgttcgg agattgatac cgtagggacg 180
gcgccttctt ctcctatctc agtcaccatg gagcccccgg agccgcatct ggtagcagac 240
gcgacccagc atcatcacca cctccaccac agccagcagc cgccgccgcc ggccgcggcc 300
ccgacgcaaa gtttgcagcc tttgccccaa cagcagcagc cgctgccgcc acagcagccg 360
ccgccgccgc ccccccagca gctgggctcg gccgcctcgg cccccaggac ttccacgtct 420
tctttttaa ttaaggacat cttgggcgac agcaaacctc tggcggcatg tgcaccctac 480
agcaccagcg tatcctctcc ccaccacacc ccgaagcagg agagcaacgc agtgcacgag 540
agcttcaggc caaagctcga gcaggaggac agcaagacca aactcgacaa gcgggaggat 600
tcccagagcg acatcaaatg ccacgggaca aaggaggaag gagaccggga gattacgagt 660
agccgtgaga gtccccctgt gagagccaag aagcctcgaa aagcaaggac agctttttcc 720
gaccaccagc tcaatcaact ggagcgtagc tttgagcggc agaagtacct gagcgtgcag 780
gatcgcatgg acctggctgc agcgctcaac ctcactgaca cccaagtcaa gacctggtac 840
cagaaccgca ggaccaagtg gaagcggcag acagcggtgg gcctggagtt gctggccgag 900
gcagggaact actcggcgct gcagaggatg tttccatcgc cttatttcta tcacccaagc 960
ctgctgggca gcatggacag cactacggcg gcggcggctg ccgctgccat gtacagcagc 1020
atgtaccgga ctcctccagc accccatccc cagctgcagc ggcccctggt gccccgtgtg 1080
ctcatccacg gcctagggcc tgggggacag ccagccctta atccattgtc cagccccatc 1140
ccaggcaccc cacacccccg gtgaaaacat tgcagcgaag gcactgcaat cccttcccca 1200
tatccctgcc caacccggac tgctgcccgt cctctccctg ctccaggcca ctaggcctcc 1260
cttgcagcca actctggaag gcagaggagt aagagaggaa gatgcttacc agtgggcagg 1320
ggacccccaa agaggagcca gccctctgct ctccatc 1357

<210> 59
<211> 2153
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7495008CB1

<400> 59
gcgcgcgtga acgcggtccc cgggaccatg ctgcggccac agcggcccgg agacttgcag 60
ctcggggcct ccctctacga gctggtgggc tacaggcagc cgccctcctc ctcctcctcc 120
tccacctcct ccacctcctc cacttcctcc tcctccacga cggccccct cctccccaag 180
gctgcgcgcg agaagccgga ggcgccggcc gagcctccag gccccgggcc cgggtcaggc 240
gcgcacccgg gcggcagcgc ccggccggac gccaaggagg agcagcagca gcagctgcgg 300
cgcaagatca acagccgcga gcggaagcgc atgcaggacc tgaacctggc catggacgcc 360
ctgcgcgagg tcatcctgcc ctactcagcg gcgcactgcc agggcgcgcc cggccgcaag 420
ctctccaaga tagccacgct gctgctcgcc cgcaactaca tcctactgct gggcagctcg 480
ctgcaggagc tgcgccgcgc gctgggcgag ggcgccgggc ccgccgcgcc gcgcctgctg 540
ctggccgggc tgcccctgct cgccgccgcg cccggctccg tgctgctggc gcccggcgcc 600
gtaggacccc ccgacgcgct gcgccccgcc aagtacctgt cgctggcgct ggacgagccg 660
ccgtgcggcc agttcgctct ccccggcggc ggcgcaggcg gccccggcct ctgcacctgc 720
gccgtgtgca agttcccgca cctggtcccg gccagcctgg gcctggccgc cgtgcaggcg 780
caattctcca agtgagggcg ggcctgggcc tggggcgcga cctcggcccg gcctcccttc 840
gctcagcttc tccgcgcccc tgctccctgc gtctgggaga gcgaggccga gcaaggaaag 900
catttcgaac cttccagtcc agaggaaggg actgtcgggc acccccttcc ccgcccccac 960
ccctgggacg ttaaagtgac cagagcggat gttcgatggc gcctcggggc agtttggggt 1020
tctgggtcgg ttccagcggc tttaggcaga aagtgctcgc tctcacccag cacatctctc 1080
tccttgtccc tggagttgcg cgcttcgcgg ggccgatgta gaacttaggg cgccttgccg 1140
tggttggcgc gccccgggtg cagcgagagg ccatccccga gcgctacctc cccggagcgg 1200
agcacgccgg ctcccagtac taggggctgc gctcgagcag tggcgggggc ggaggggtgg 1260
ttcttttcct tctcctccgc cagaggccac gggcgccctt gttcccgccg gccaggtcct 1320
atcaaaggag gctgccggaa ctcaagaggc agaaaaagac cagttaggcg gtgcagacgg 1380
tctgggacgt ggcagacgga cggaccctcg gcggacaggt ggtcggcgtc ggggtgcggt 1440
gggtaggggc gaggacaacg cagggtgcgc tgggttggga cgtgggtcca cttttgtaga 1500
ccagctgttt ggagagctgt atttaagact cgcgtatcca gtgttttgtc gcagagagtt 1560
ttcgctctta aatcctgggg gtttcttaga aagcaactta gaactcgaga ttcacctttc 1620
gtttcccttt ccccaaaagt agcgtaacca acatttaagc ttgcttaaaa acgaaaacca 1680
accgccttgc atccagtgtt cccgatttac taaaataggt aaccaggcgt ctcacagtcg 1740
ccgtcctgtc aagagcgcta atgaacgttc tcattaacac gcaggagtac cgggagccct 1800
gaaccgcccg ctgctcggcg gatcccagct gcggtggcga cggcgggaag gcgctttccg 1860
ctgttcctca gcgggccggg cccttgacca gcgcggcccg caggtcttcc ttctcgccgt 1920
cttgcagttg aagagctaca tacgtagtca gtttcgattt gttacagacg ttaacaaatt 1980
cctttaccca aggttatgct atgacctttc cgcagtttac tttgattttc tatgtttaag 2040
gttttggttg ttggtagtag ccgaatttaa ctggcacttt attttacttc taaccttgtt 2100
tcctgacggt gtacagaatc aacaaaataa aacatttaaa gtctgatttt tta 2153

<210> 60
<211> 1104
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7073515CB1

<400> 60
atgttcggga aaccagacaa aatggacgtt cgatgccact cggacgccga ggctgcccgg 60
gtctcgaaga acgcgcacaa ggagagtcgg gagagcaagg gcgcggaggg gaacctccca 120
gccgccttcc tcaaggagcc gcagggcgcc ttctcagcgt cgggcgctgc tgaggattgt 180
aacaaaagta aatccaattc cgcagcggac ccggattact gccgccggat cctggtccga 240
gatgccaagg ggtccatccg agagatcatc ctgcccaagg gcctggactt ggaccggcct 300
aagaggacgc gcacgtcctt caccgcggag cagctctatc ggctggagat ggagttccag 360

cgctgccagt acgtggtggg ccgcgagagg accgagctcg cccggcagct taacctctcc 420
gagacccagg tgaaggtctg gttccagaac cggcgcacca agcagaagaa ggaccagggc 480
aaggactcgg agctacgctc ggtggtgtcg gagaccgcgg ccacgtgcag cgtgctacgg 540
ctgctggagc agggccgcct gttgtcgccg cccggcctgc ctgcgctgct gccgccttgc 600
gccacgggcg ctctcggttc agcgctgcgc gggcccagct tgccggccct gggcgcgggc 660
gccgctgcag gctcggccgc cgcagccgcc gccgccgccc cgggcccagc gggcgctgca 720
tccccgcacc cgccggctgt gggcggtgct ccaggtcccg ggcccgccgg gccgggggga 780
ttgcacgcat gcgccccggc cgcgggccac agcctcttca gcctgccggt gccctcgctg 840
ctcggctccg tcgccagccg cctgtcctcc gccccgttaa caatggctgg ttcgctagct 900
gggaatttgc aagaactctc cgcccgatat ctgagctcct cggccttcga gccttactcc 960
cggaccaaca ataaagaagg ggccgagaaa aaagcgctgg actgatttta agtgtttccc 1020
tgtatttata tttatagtat ctattgtggt gatatttatg gactcacgcg cattaggtgc 1080
ccagagctcc tgcgctgggc tcct 1104

<210> 61
<211> 2597
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3356640CB1

<400> 61
gcggggtctt tgtctcgctg cagcgggtgc tgcaggtctg gccttcactt ttctgcgtcc 60
tcttactcct agaggcccag cctctgtggc gctgtgatct ggttatcggg agattcacag 120
ctaagacgcc aggatccccc ggaagcctag aaatgggacc actgacattt atggatgtgg 180
ccatagagtt ctctttggag gagtggcagt gcctggacac tgcacagcgg aatgtatata 240
ggcatgtgat gttagagaac tacagaaacc tggttttctt ggccagacct gattacctgt 300
ctggaacaag gaaaagagcc ctggaatatg aagagacatg agatggtggt agccaaacat 360
tcagctctgt gttctcgttt tgcccaagac ctttggctag agcagaacat aaaagattct 420
ttccaaaaag tgacactgag cagatatgga aaatatggac ataagaattt acagttaaga 480
aaaggctgta aaagtgtgga tgagtgtaag gaacaccaag gaggttataa tggacttaac 540
caatgtttga aaattaccac aagcaaaata tttcaatgta ataaatatgt aaaagtcatg 600
cataaatttt caaattcaaa tagacacaag ataagacata ctgaaaataa acatttcaga 660
tgtaaagaat gtgacaaatc actttgcatg ctttcacgcc taactcaaca taaaaaaatt 720
catactagag agaatttcta caaatgtgaa gagtgtggaa aaacctttaa ctggtccaca 780
aacctttcta aacctaagaa aattcatact ggagaaaaac cctacaaatg tgaagtatgt 840
ggaaaagcct ttcaccaatc ctcaatcctt actaaacata agataattcg tactggagaa 900
aaaccctata aatgtgcaca ctgtggcaaa gcctttaaac agtcctcaca ccttactaga 960
cataagataa ttcatactga agagaaaccc tacaaatgtg aacaatgtgg caaggtcttt 1020
aagcagtccc caacccttac taaacatcag ataatttata ctggagagga accatacaaa 1080
tgtgaggaat gtggcaaagc ttttaaccta tcttaacaac ttactgaaca taagaaaatt 1140
tacactagag agaaagccta caaatgtgaa gaatgtggca aagcctttaa ccagttttca 1200
acccttatta cacataagat aattcatagc ggagagaaac cccacaaatg tgaagaatgt 1260
ggcagagctt ttaaccagtc cgcaaagctc actgaacata agttaattca tactggagaa 1320
aaaccctaca aatgtaaaga atgtggaaaa gcttttcacc gatactcaat ccttagtaca 1380
cataagaaaa ttcatactgg ggagaaaccc cacaaatgtg gagaatgcgg aaaagccttt 1440
aactggtcct caactcttat tacacataag ataattcaca gtggagaaaa accctacaaa 1500
tatgaagaat gtggcaaagc ttttaaccag tcctcacacc ttatgagaca taagaaaatt 1560
catagtaaag agaaacctta caaatgtgaa cagtgtggca aggtctttaa gaagtcctca 1620
actcttactg cacataagat cattcatact ggagagaaac cttacaaatg tgaggaatgt 1680
ggcaaaggtt ttagccaact ctcaaacctt actaaacaca agaagattca tactagagag 1740
aaaccctaca aatgtgaaga atgtggcata tcttttaacc agttctcaca acttgctata 1800
cataagatga ttcacacttg aatgaaaccc tacaaatgtg aacgatgtgg cagttgtttt 1860
aactagttct cgaactttac tatgcataag aaaattcaaa ctggagagaa actctacaaa 1920
tgtgaagaat gtggcaaagc ttttaaccaa gtctcaacac ttactataca taagataatt 1980
tatactggag caaaaccttg gaaattcaaa gaatgtggta aaacttataa tcctcaaaac 2040
ttcttacacc taaaattcat gcaggagaga aaccccacaa atgtgaaaaa tttggtaaat 2100
tctttaacaa gtcttcaacc ctttctgcac ataatataat tcatactgga gagaaacccc 2160
acaaatatga agaatgtggt aatgctttta accaattctc aaatcttact aaacaaaatt 2220
aatactgaaa atgttacaaa ccagaaaaat gtgaaaatga ttttaacaaa accttcaaat 2280
ttttctaaac ataaaggaaa tcatactggt aagaaattat aaaaatgtga agaatgtgac 2340

aaagccttta aatggttgtc acacttgatt gtaggtaaga taattcatac tggcagaaac 2400
tcccagaagt gtgaagaata tggcaaaact ttaattccta taccttattg cacaggaaag 2460
catttatact tcagaaaatg ttgtactgat ataaagaatg tagaaaagcc attaatatgt 2520
gcttacatct tattcaacat tagagagtta gtacttaata aaagcattat aaatgcaatt 2580
actgtcaaaa aaaaaaa 2597

<210> 62
<211> 1959
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2015706CB1

<400> 62
gggacgttga gtccccacag acctggaaat tgcggcccct ctttctcaac ccagagcaaa 60
ttgagacgtc cgggctggag ggcagttgtg cgatcttgat tcacttcaac ctctgctctc 120
cgggctcaag caattctccc acctcagcct gagtctcact ctgtccttca ggctgtgagt 180
gcagtctcct gatcttggct cactgcaacc tctgtctcct gggttcaagt gattcttgtg 240
cctcagccta ccaagtagct gagatacagg attgacttct aaagactctt ggtacctgag 300
gaagaaaccc agaagaggaa gaggaaagca aaggagtcgg ggatggctct ttctcagggt 360
ctactgacat tcagggatgt ggccatagaa ttctctcagg aggagtggaa atgcctggac 420
cctgctcaga ggactctata cagagacgtg atgctggaga attataggaa cctggtctcc 480
ctggatatct cttccagatg catgatgaat acattgtcat caacagggca aggcaataca 540
gaagtgatcc acacagggac attgcagaga caagcaagtt atcacattgg agcattttgc 600
tcccaggaaa ttgagaaaga cattcatgac tttgtgtttc agtggcaaga agatgaaaca 660
aatgaccatg aagcacccat gacagaaata aaaaagttga ctagtagtac agaccgatat 720
gatcaaaggc atgctggaaa caagcctatt aaaggtcagc ttgaatcaag atttcatttg 780
catttgcgaa gacataggag aattcatact ggagagaaac cttacaaatg tgaagaatgt 840
gagaaagttt tcagttgcaa atcacatctt gaaatacata ggataattca tactggagag 900
aaaccataca aatgtaaggt ttgcgacaag gcttttaagc atgattcaca cctggcaaaa 960
catactagaa ttcacagggg agacaaacat tacacatgta atgaatgtgg caaggttttt 1020
gatcaaaaag caacccttgc atgtcatcat agaagtcata ctggagagaa accttataag 1080
tgtaatgagt gtggcaagac ctttagtcag acatcacacc ttgtgtacca tcatagactg 1140
catactggag agaaacctta caaatgtaat gagtgtggca agaccttcgc tcgaaattca 1200
gtcctcgtaa ttcataaggc agttcatact gcagagaaac cctataagtg taatgaatgt 1260
ggcaaggttt ttaagcaacg agcaactctt gcaggacatc gtagagttca cactggagag 1320
aaaccttaca gatgtgaaga atgcgacaaa gttttcagtc gcaaatcaca tcttgaaaga 1380
cataggagga ttcatactgg agagaaacca tacaaatgta aggtttgtga caaggctttc 1440
cggagtgatt cacgtcttgc agaacatcag agagttcata ctggagagag accttacaca 1500
tgtaatgaat gtggcaaggt ttttagtaca aaagcatacc tcgcatgtca tcaaaaactt 1560
catactggag agaaacttta cgaatgtgaa gaatgtgaca aagtttacat tcgcaaatca 1620
caccttgaaa gacataggag gattcataca ggagagaaac ctcacaagtg tggtgattgt 1680
ggtaaagcct ttaattcacc ttcacacctt attaggcatc agagaatcca tactggacag 1740
aaatcttaca aatgtcatca gtgtggcaag gtctttagtc tgaggtcact ccttgcagaa 1800
catcagaaaa ttccttttgg agacaattgt ttcaagtgca atgagtatag caaaccatca 1860
agcattaatt gacatcagag tcaaatcagc attgacctga gtttgagttg acttaacatt 1920
gatttcaagc attaatcgac attaaactgt ttatgttaa 1959

<210> 63
<211> 1401
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6920755CB1

<400> 63
attttaacta aaggttatta tcataaagca ggtgtttgct gaagacagct tactcagatc 60
actactgcct ggaggtggtt gatatatcct ggtgtaaacc ttcaagaagg gcacaggcag 120
gaaaacatga gccagcaact gaagaaacgg gcaaagacaa gacaccagaa aggcctgggt 180

ggaagagccc ccagtggggc taagcccagg caaggcaagt caagccaaga cctgcaggcg 240
gaaatagaac ctgtcagcgc ggtgtgggcc ttatgtgatg gctatgtgtg ctatgagcct 300
ggccctcagg ctctcggagg ggatgatttc tcagactgtt acatagaatg cgtcataagg 360
ggtgagtttt ctcaacccat cctggaagag gactcacttt ttgagtcctt ggaataccta 420
aagaaaggat cagaacaaca gctttctcaa aaggttttcg aagcaagctc ccttgaatgt 480
tctttggaat acatgaaaaa aggggtaaag aaagagcttc cacaaaagat agttggagag 540
aattcgcttg agtattctga gtacatgaca ggcaagaagc ttccgcctgg aggaatacct 600
ggcattgacc tatcagatcc taaacagctc gcagaatttg ctagaaagaa gccccccata 660
aataaagaat atgacagtct gagcgcaatc gcttgtcctc agagtggatg cactaggaag 720
ttgagggata gagctgccct gagaaagcat ctcctcattc atggtccccg agaccacgtc 780
tgtgcggaat gtgggaaagc gttcgttgag agctcaaaac taaagagaca tttcctggtt 840
catactggag agaagccgtt tcggtgcact tttgaaggat gcggaaagcg cttctctctg 900
gactttaatt tgcgtacgca cgtgcgcatc cacacggggg agaaacgttt cgtgtgtccc 960
tttcaaggct gcaacaggag gtttattcag tcaaataacc tgaaagccca catcctaacg 1020
catgcaaata cgaacaagaa tgaacaagag ggaaagtagt cctccaacag gatgaagcag 1080
attaacagaa gagtgatcag tgacaaacat gcctcattga ttattgtttc taggaaggaa 1140
tttctaaatc aatattgcaa ccccaaaagc ggttataatt tggtgttact aagatgctcc 1200
tacactttgt gataccgttt taaggacatg gtgcattttt ttttctttta tttgttttat 1260
ttagaacttt ttttatttgt tttatttaga actttgtgtg ttcttaaagt gtgcttccaa 1320
caggaaggtc agtgataaat tttcaaaagc ataaccttca atatattatc tgttggatta 1380
ttggatataa gacttatttt c 1401

<210> 64
<211> 3406
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 444179CB1

<400> 64
cactgtgtct ctgttgacct gccagatagg aaatagcgcc agagttgctg caatccacgt 60
ctccctggaa ttcaggagca gcttggctgg aaggctctgg ctggaaactt gcagaaggct 120
gtcagctggg tctgcatcat ctgaaggctc aacgggtacc gggggccact ccatggcctg 180
cgagtcagag ctggccgtcg gtgggaggcc tcagttcctc tccacatagg cctctccaca 240
gggccacttg agcatcttcc caccatggca actggtgtcc tctagagcaa catccaggag 300
accaaggcgg aagcagctgt agtgccttct aggacctggt cctgaaagtc acccctgcac 360
ccacccccgc tgtctcctca gccatattct gctggtcaca cagagcgctg cgggataaag 420
gaggagcgtc ctgcttcccg gctgccctgt tgctgtcgga gtcacaggat ggcggctgtc 480
atcctgccct cgactgctgc tccgtcttcc ctgttcccag cctctcagca aaaaggacac 540
acacagggcg gagagctggt taatgagctc ctgacaagct ggctacgggg cttggtaacc 600
ttcgaggatg tggccgtgga gttcacccag gaggagtggg cgttgctgga ccctgcccaa 660
aggacactgt acagggatgt gatgctggag aactgcagga acctggcctc actagggtgt 720
cgtgttaata aacccagtct gatatcccag ttggaacaag acaagaaggt ggtgacagag 780
gaaagaggaa ttctaccaag cacctgtcca gatttggaga ctctacttaa agccaaatgg 840
ttaactccta agaagaatgt tttcagaaaa gaacagtcta aaggtgtaaa aacggaaaga 900
agtcatcgtg gagtgaaact caatgaatgt aatcagtgtt ttaaagtctt cagcacgaaa 960
tctaacctaa ctcagcacaa gagaattcat actggagaaa aaccctatga ctgtagtcaa 1020
tgtgggaagt ccttcagtag cagatcttac cttactattc ataagagaat ccataatggg 1080
gagaaaccct atgaatgcaa tcactgtggg aaagcattta gtgatccctc atcccttaga 1140
ctgcatttga gaattcacac tggagaaaaa ccctatgaat gtaaccagtg ttttcacgtt 1200
ttccgcacca gttgtaacct caaaagccac aagaggattc acacggggga gaatcaccat 1260
gaatgtaatc agtgtggaaa agctttcagc acaaggtcct ctctcactgg gcacaatagc 1320
attcatacag gggagaaacc ttatgaatgt cacgattgtg ggaaaacctt caggaagagc 1380
tcctatctga cacagcacgt aagaactcat actggagaaa aaccctatga atgtaacgag 1440
tgtgggaaat ccttcagcag tagcttttct cttactgtgc acaagagaat acataccgga 1500
gagaaaccct acgagtgcag tgactgtgga aaagccttta ataatctctc agctgtgaag 1560
aaacacttaa gaactcacac tggagaaaaa ccctatgaat gtaatcattg tgggaaatcc 1620
ttcacaagta actcctatct ttctgtgcac aagagaatac ataatagatg gatatgaatt 1680
actgcaggaa cttctggagg aaagcactca ttgatctttc atccctaaga tagtttgaga 1740
gagctcacac tggatatata agttatttgt tgcagcataa caaatgatcc ccgcaaattt 1800
tgtggcttca aacacgaaat gtttattctc taacaatttc tacagatcag gaattcaaga 1860

acagcttatc tctgtagtta cactgcagga tctctcatgc ctttacttac aatcaagatg 1920
tcagccaaag ctatagtcat ctaacatctt tactgatgaa cagtccactt ccaagatgac 1980
ttattcacat gcctggaaag ttgatgttgg ttgttgtcag gggaccttca gtcattacca 2040
tgtaggcctc tttacaaggc ccagctcctt gacaatttgg caataagttc tcgccaaggt 2100
tagtgatcca agagaggacc aatacagagg ccaaaatggc ttttatgacc tatcctcagg 2160
aggactgcta agctttcact tatttcctat tgtgttggtc tcacaacaca ccaatcctga 2220
tacaatgtga gagactgcag aaggtatgaa taccaggaaa cagataactc agaagcatct 2280
gggaaatttg ccatggaatt aatgaagaaa agccctcagc acatcatcct tctactctaa 2340
tcaatatgaa atagcattca gtaactccta ttaattcact ctgaagagaa acctttgag 2400
gacaatctgt atggtaaagc tctcagctcg aattctcacc ttcatgggcc cagaagatta 2460
tgtactgaga gaatcctgaa gaatggaaca gctgtagaaa gccttcagtg ctatcagcag 2520
ggattcatgt gacagctcac actaagggag aaaacctgtg aaggtctcag aaggcttcag 2580
tgacagtcat cccttaagac gcactggaaa tcacacaatg gaacaacact cagaaatgcg 2640
gaataccctg cagcaaaagt gttcacatta ctgagcaaac tatagtgtgg tattctgcag 2700
tgactatggg aaagccttga atgttctatc ggtttttaag ggacttgaga attaattctg 2760
gagagaatgc cccattgaac atcatcaata ttggagagct ttctgttttt ctacatttgt 2820
taggaaactt gtgagcattc acactacaga gaaacttgaa atataaagaa gaagagaaag 2880
ccttcagtga tgcctctgtg ttagggaaaa tatggaactt ctccctggat gcaaaaccta 2940
tgagtatatt aatattggaa aatttttcag tgattcttct ctttcttgta tatgagagaa 3000
cttatatgga gaaaccccta ggaatgtaat cagtgttagg atgcctcagc ctgaactctt 3060
cactgagtgg ccacaatttt cactgggaac aaaaagtata atcactgttt tgagtgtggg 3120
atatccttta tcagtgtctc atctgtagat tggactgctg gctcattaat ttttttagtc 3180
tttttttctt ttaatataaa catttgtgta tagctgttcc ctaaaataaa cattaacata 3240
tttcataatt ttaatgcaat gtatttatta taattaattt gcttgttaaa gacatccaca 3300
cattgcatat tcaaaaagtt atttccaaat accttctgag tgattcagtt tatcataatg 3360
gaaattagta tttataaaca cattttcta atgtagtggt attttt 3406

<210> 65
<211> 2718
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5628380CB1

<400> 65
caagaattag agacaagcgg tcagcagagc ctcagtgctg atcgtcggag cttgggagca 60
ggaagatgtc gaatgaactt gatttcaggt ctgtgcggct gctaaagaac gacccagtca 120
acttgcagaa attctcttac actagtgagg atgaggcctg gaagacgtac ctagaaaacc 180
cgttgacagc tgccacaaag gccatgatga gagtcaatgg agatgatgac agtgttgcgg 240
ccttgagctt cctctatgat tactacatgg gtcccaagga gaagcggata ttgtcctcca 300
gcactggggg caggaatgac caaggaaaga ggtactacca tggcatggaa tatgagacgg 360
acctcactcc ccttgaaagc cccacacacc tcatgaaatt cctgacagag aacgtgtctg 420
gaaccccaga gtacccagat ttgctcaaga agaataacct gatgagcttg gaggggggcct 480
tgcccacccc tggcaaggca gctcccctcc ctgcaggccc cagcaagctg gaggccggct 540
ctgtggacag ctacctgtta cccaccactg atatgtatga taatggctcc ctcaactcct 600
tgtttgagag cattcatggg gtgccgccca cacagcgctg gcagccagac agcaccttca 660
aagatgaccc acaggagtcg atgctcttcc cagatatcct gaaaacctcc ccggaacccc 720
catgtccaga ggactacccc agcctcaaaa gtgactttga atacaccctg ggctccccca 780
aagccatcca catcaagtca ggcgagtcac ccatggccta cctcaacaaa ggccagttct 840
accccgtcac cctgcggacc ccagcaggtg gcaaaggcct tgccttgtcc tccaacaaag 900
tcaagagtgt ggtgatggtt gtcttcgaca atgagaaggt cccagtagag cagctgcgct 960
tctggaagca ctggcattcc cggcaaccca ctgccaagca gcgggtcatt gacgtggctg 1020
actgcaaaga aaacttcaac actgtggagc acattgagga ggtggcctat aatgcactgt 1080
cctttgtgtg gaacgtgaat gaagaggcca aggtgttcat cggcgtaaac tgtctgagca 1140
cagacttttc ctcacaaaag ggggtgaagg gtgtcccccct gaacctgcag attgacacct 1200
atgactgtgg cttgggcact gagcgcctgg tacaccgtgc tgtctgccag atcaagatct 1260
tctgtgacaa gggagctgag aggaagatgc gcgatgacga gcggaagcag ttccggagga 1320
aggtcaagtg ccctgactcc agcaacagtg gcgtcaaggg ctgcctgctg tcgggcttca 1380
ggggcaatga gacgacctac cttcggccag agactgacct ggagacgcca cccgtgctgt 1440
tcatccccaa tgtgcacttc tccagcctgc agcgctctgg aggggcagcc ccctcggcag 1500
gacccagcag ctccaacagg ctgcctctga agcgtacctg ctcgcccttc actgaggagt 1560

ttgagcctct gccctccaag caggccaagg aaggcgacct tcagagagtt ctgctgtatg 1620
tgcggaggga gactgaggag gtgtttgacg cgctcatgtt gaagacccca gacctgaagg 1680
ggctgaggaa tgcgatctct gagaagtatg ggttccctga agagaacatt tacaaagtct 1740
acaagaaatg caagcgagga atcttagtca acatggacaa caacatcatt cagcattaca 1800
gcaaccacgt cgccttcctg ctggacatgg gggagctgga cggcaaaatt cagatcatcc 1860
ttaaggagct gtaaggcctc tcgagcatcc aaaccctcac gacctgcaag gggccagcag 1920
ggacgtggcc ccacgccaca cacaacctct ccacatgcct cagcgctgtt acttgaatgc 1980
cttccctgag ggaagaggcc cttgagtcac agacccacag acgtcagggc cagggagaga 2040
cctaggggct cccctggcct ggatccccat ggtatgcttg aatctgctcc ctgaacttcc 2100
tgccagtgcc tccccgtacc ccaaaacaat gtcaccatgg ttaccaccta cccagaagac 2160
tgttccctcc tcccaagacc cttgtctgca gtggtgctcc tgcaggctgc ccgttaagat 2220
ggtggcggca cacgctccct cccgcagcac cacgccagct ggtgcggccc ccactctctg 2280
tcttccttca acttcagaca aaggatttct caacctttgg tcagttaact tgaaaactct 2340
tgattttcag tgcaaatgac ttttaaaaga cactatattg gagtctcttt ctcagacttc 2400
ctcagcgcag gatgtaaata gcactaacga tcgactggaa caaagtgacc gctgtgtaaa 2460
actactgcct tgccactcac tgttgtatac atttcttatt tacgattttc atttgttata 2520
tatatatata aatatactgt atatatatgc aacattttat atttttcatg gatatgtttt 2580
tatcatttca aaaaatgtgt atttcacatt tcttggactt tttttagctg ttattcagtg 2640
atgcattttg tatactcacg tggtatttag taataaaaat ctatctatgt attacgtcac 2700
attacttcaa aaaaaaaa 2718

<210> 66
<211> 3325
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7493789CB1

<400> 66
gccgagcccg ccgcatgtgg cccggctccc ggacacctcc ccggcgtcct ccgcgccggc 60
cgctcctgcc ccgacgtcgc tccggcacgg ctcggggccc agaggcgagg cgaggacgcc 120
gggcaagcca ggcagcggaa ctgacgccgg cgagcttccg gggcggcccc gggcaggtcg 180
gcggcggcgg cccgcagtcg tggaggagcg gtgggagcgt cggcggccgc gggcgatgca 240
acttccggac gggactcccc tctgtccgcg cctcacatct ccccttcctc tcgcctagtc 300
ctgtgccgtt ttccgtccgc gactcttccg gcccagagct ttcggagtgc ggttgctcag 360
gggaagccgt cgccgccccc gcctcggggc cgagtgagag tgcccgtcgc gtcgcgccgc 420
gtcgcccccc gggccgcctc cttgccgcca gtggcgggct ccgttctccc tcgaagcact 480
ccccccagct ccatgaatgg aaatcggctc cgcaggaccc gctggggccc agcccctact 540
catggtgccc agaagacctg gctatggcac catgggcaaa cccattaaac tgctggctaa 600
ctgttttcaa gttgaaatcc caaagattga tgtctacctc tatgaggtag atattaaacc 660
agacaagtgt cctaggagag tgaacaggga ggtggttgac tcaatggttc agcattttaa 720
agtaactata tttggagacc gtagaccagt ttatgatgga aaaagaagtc tttacaccgc 780
caatccactt cctgtggcaa ctacaggggt agatttagac gttactttac ctggggaagg 840
tggaaaagat cgacctttca aggtgtcaat caaatttgtc tctcggtaca cacctgtggg 900
gcgttcattt ttctccgctc cagaaggata tgaccaccct ctgggagggg gcagggaagt 960
gtggtttgga ttccatcagt ctgttcggcc tgccatgtgg aaaatgatgc ttaatatcga 1020
tgtttctgcc actgccttct acaaagcaca acctgtaatt cagttcatgt gtgaagttct 1080
tgatattcat aatattgatg agcaaccaag acctctgact gattctcatc gggtaaaatt 1140
caccaaagag ataaaaggtt tgaaggttga agtgactcat tgtggaacaa tgagacggaa 1200
ataccgtgtt tgtaatgtaa caaggaggcc tgccagtcat caaacctttc ctttacagtt 1260
agaaaacggc caaactgtgg agagaacagt agcgcagtat ttcagagaaa agtatactct 1320
tcagctgaag tacccgcacc ttccctgtct gcaagtcggg caggaacaga aacacaccta 1380
cctgccacta gaagtctgta atattgtggc agggcaacga tgtatcaaga agctaacaga 1440
caatcagact tccactatga tcaaggcaac agcaagatct gcaccagata gacaagagga 1500
aattagcaga ttggtaagaa gtgcaaatta tgaaacagat ccatttgttc aggagtttca 1560
atttaaagtt cgggatgaaa tggctcatgt aactggacgc gtacttccag cacctatgct 1620
ccagtatgga ggacggaatc ggacagtagc aacaccgagc catggagtat gggacatgcg 1680
agggaaacaa ttccacacag gagttgaaat caaaatgtgg gctatcgctt gttttgccac 1740
acagaggcag tgcagagaag aaatattgaa gggtttcaca gaccagctgc gtaagatttc 1800
taaggatgca gggatgccca tccagggcca gccatgcttc tgcaaatatg cacaggggc 1860
agacagcgta gagcccatgt tccggcatct caagaacaca tattctggcc tacagcttat 1920

tatcgtcatc ctgccgggga agacaccagt gtatgcggaa gtgaaacgtg taggagacac 1980
acttttgggt atggctacac aatgtgttca agtcaagaat gtaataaaaa catctcctca 2040
aactctgtca aacttgtgcc taaagataaa tgttaaactc ggagggatca ataatattct 2100
tgtacctcat caaagacctt ctgtgttcca gcaaccagtg atcttttttgg gagccgatgt 2160
cactcatcca cctgctggtg atggaaagaa gccttctatt gctgctgttg taggtagtat 2220
ggatgcacac ccaagcagat actgtgccac agtaagagtt cagagacccc gacaggagat 2280
catccaggac ttggcctcca tggtccggga acttcttatt caattttata agtcaactcg 2340
gttcaagcct actcgtatca tcttttatcg ggatggtgtt tcagaggggc agtttaggca 2400
ggtattatat tatgaactac tagcaattcg agaagcctgc atcagtttgg agaaagacta 2460
tcaacctgga ataacctaca ttgtagttca gaagagacat cacactcgat tattttgtgc 2520
tgataggaca gaaagggttg gaagaagtgg caatatccca gctggaacaa cagttgatac 2580
agacattaca cacccatatg agttcgattt ttacctctgt agccatgctg gaatacaggg 2640
taccagtcgt ccttcacact atcatgtttt atgggatgat aactgcttta ctgcagatga 2700
acttcagctg ctaacttacc agctctgcca cacttacgta cgctgtacac gatctgtttc 2760
tatacctgca ccagcgtatt atgctcacct ggtagcattt agagccagat atcatcttgt 2820
ggacaaagaa catgacagtg ctgaaggaag tcacgtttca ggacaaagca atgggcgaga 2880
tccacaagct cttgccaagg ctgtacagat tcaccaagat accttacgca caatgtactt 2940
cgcttaaata gtccaagtat attctctgag aggaagtact gaaagatgaa ttgacataca 3000
acgtatgttt ccagtgaagt caattgagta aggacacctc cagccataca gaaaccaaca 3060
ctgtgtgggg gccaaggtct gatccttatg ttaatacaag gaagattgtt tacttcatca 3120
aggaacacag catcattatg caatatgaaa ccagccaact gctttttgtg cggtctccta 3180
taggaagtat cgcaattgtt ttgttttcat ttcttgtagt ctaacccttt taatgccttt 3240
acctcaagtt gcttggcagc acaactatct ttgcaaaaaa aagtaaagaa aaagtaaatg 3300
atggtttaaa aaatacacac aaaaa 3325

<210> 67
<211> 8114
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2075194CB1

<400> 67
cgcctcttcc tcccaaaggg aagacggtga gccggaggag tcagtcagag gggcgagcag 60
gagcgattcc gtcgccaaac aggttatgag tgccagtgag ccgccttaga tagaagcatc 120
gtcagcactt tattaatgat ggatagtgag aataaacccg aaaatgacga ggatgaaaag 180
ataaacaaag aagcacaaga cttgacaaag ctttcatccc ataatgaaga cggtgggcct 240
gtatctgatg tgatagcaag tttccctgag aattctatgg gcaaaagagg tttttcagaa 300
tcatcgaact ctgatagtgt tgttatagga gaagacagaa ataaacatgc ttccaaacgc 360
aggaaattag atgaggcaga gccccttaaa tctggaaagc aaggtatttg tagattagaa 420
acttctgaga gctcagtcac agaaggggat attgcattag atgaaacagg gaaggagacc 480
tttctgagtg actgcacagt tggaggcaca tgtctcccaa atgccctctc cccttcttgc 540
aattttagca ctattgatgt tgtttctctg aaaacagaca ctgaaaaaac atctgctcag 600
gaaatggttt cccttgatct ggaaagagaa tctcctttcc ccccgaaaga aattagtgtt 660
agttgtacca ttgggaatgt agatacagtt ctcaaatgcc agatctgtgg gcatttgttt 720
tcttcttgct ctgacttgga aaaacatgct gagtctcaca tgcagcagcc taaggaacat 780
acctgttgtc actgcagcca caaagcagag agcagctcag cactacatat gcatatcaaa 840
caagcacatg ggccacagaa ggtcttttcc tgtgatcttt gtggttttca gtgttcagaa 900
gaaaacttgt tgaatgcaca ttatcttggc aaaacacatc tccgtcgtca gaatctggct 960
gctcgtggag gatttgtaca gatcttaaca aaacaacctt ttcctaaaaa accacgtaca 1020
atggcaacaa aaaatgttca ctcaaaacca agaacttcta aatcaatagc aaagaatagt 1080
gattcaaaag gattacgaaa tgtgggaagc acgtttaaag atttcagagg aagtatttct 1140
aaacaaagtg gtagtagcag tgagcttctt gttgaaatga tgccttccag aaatactttg 1200
tcacaggaag tagagattgt tgaagaacat gttacttccc ttggtctagc tcagaatcct 1260
gaaaaccaga gtagaaagct agacacctta gtaacctcag agggtctctt agagaaattg 1320
gaatctacaa aaaataccct tcaggcagca cacggtaaca gtgtaacctc gaggccaaga 1380
cctgagcgaa atattcttgt gttgggtaat agctttcgtc gacgaagcag cactttcacc 1440
ttgaagggcc aggcaaagaa aaggtttaat cttttaggaa ttaaaagagg tacaagtgaa 1500
actcagagga tgtatatgaa acacttgaga acacagatga aaacacacga tgcagaatca 1560
gtgctgaaac acctggaagc gtgcagcagt gtgcagagag tgtgtgtgac tacctcagaa 1620
acccaggagg cagagcaggg ccaggggagt gcccgtcctc cggactccgg gctgcattcc 1680

ctgacagtga agccagcttc tggctctcag acgttgtgtg cttgtacaga ctgtgggcaa 1740
gtagctacaa ataggacaga tttggaaatc catgtgaaaa ggtgccatgc cagagagatg 1800
aaattttact gccgtacttg tgacttctct agtatgtcaa gaagggactt agatgaacat 1860
ttgcacagta accagcatca gcaaactgct tctgtcctga gttgtcagtg ttgttcattt 1920
atatccttgg atgaaataaa tcttagagac cacatgaagg aaaagcacaa tatgcatttt 1980
ctttgcaccc cttgtaatct gttctttttg tctgaaaaag atgtggaaga acacaaagcc 2040
accgagaagc atattaattc attggttcaa ccaaagactt tgcaatcatc taacagtgat 2100
ttggttttac agactttacc tttgagtact ttagaatcag aaaacgcaaa agagtctatg 2160
gatgactcag gaaaagcatc tcaggaagaa cctctgaagt ccagggtaag ccatggtaat 2220
gaagtgaggc attccagtaa gcctcagttt cagtgtaaga agtgttttta taaaacaaga 2280
tcttctactg ttctcacgag acatataaag cttcggcatg gtcaagacta tcattttctt 2340
tgtaaagctt gtaatcttta ttcattgagc aaagaaggaa tggagaaaca cattaaaaga 2400
agcaagcatc ttgaaaatgc taagaaaaat aatattggct taagctttga agaatgtatt 2460
gaaagggtat gtataggtgc aaatgataaa aaagaagagt ttgatgtttc cggaaatgga 2520
aggattgaag gccatatagg tgtgcaatta caagagcatt cctatcttga gaagggcatg 2580
ctggcgtctg aggaactgtc acagtctggt ggtagcacca aagatgatga attagcttca 2640
accactactc caaagagagg gagacctaaa ggtaacatct cacggacgtg ttcacactgt 2700
ggccttttgg cctctagtat tacaaacttg actgttcaca ttagacgaaa acacagtcac 2760
cagtatagtt atttatgcaa agtgtgtaag tattacactg taactaaggg agatatggaa 2820
cgtcattgtg ccaccaagaa acataaagga cgggtagaaa tagaagcaag tggaaaacac 2880
agttcagata tcattgttgg ccctgaaggg ggtagccttg aagctggtaa aaagaatgct 2940
ggctcagcag tgaccatgtc agatgaacat gctaacaaac cagctgagtc acccacctcc 3000
gttttagaga agccagatcg tggaaactca attgaagctg aagttgaaaa tgtatttcat 3060
tctctagatg gagaagttaa cagccatctt cttgataaaa aggagcaaat atcttcagag 3120
ccagaggact tcgcccagcc gggggatgtg tactcccaga gagatgttac aggcacaggt 3180
gagaataagt gtttgcactg tgagtttagt gctcactcct ctgcttctct agagctgcat 3240
gtaaaacgga aacatacaaa agagtttgag ttttattgca tggcatgcga ttactacgcg 3300
gtgactcgtc gcgagatgac caggcatgca gcaacagaga agcacaaaat gaaaaggcag 3360
tcttatctca actctgctaa tgtagaagct ggttctgcag acatgtccaa aaacatcatt 3420
atgcctgaag aagagcatca acaaaattct gaggaatttc aaataatttc aggtcaacca 3480
tctgatactc ttaaatctag aaatgctgca gattgctcta ttttaaatga gaatactaat 3540
ttagatatgt ctaaagtgct ctgcgctgct gactctgtag aagttgagac tgaagaagaa 3600
tctaatttca atgaagacca ttccttttgt gagactttcc aacaggctcc tgtcaaggat 3660
aaagttagga aacctgagga gatgatgtca cttactatgt cctcaaacta tggctcccca 3720
agcagatttc aaaatgaaaa ttcaggaagc tctgccttaa attgtgagac agcaaagaaa 3780
aaccatgaga tatcgaatga tgcaggtgag ctgcgtgtcc attgtgaggg tgaaggagga 3840
aacgcaggag acggtggagg tgttgtcccc cacagacacc tgtgccctgt gacgctcgat 3900
ggggagcgct cggctgaaag ccctgtgctc gttgtgacaa gaataaccag agaacaggga 3960
aatctggaga gcgggggtca gaacagagtt gcacgtgggc atggtttgga agacttgaaa 4020
ggtgtccaag aagatcccgt tctggggaat aaggaaattc tgatgaattc acaacatgaa 4080
acagaattta ttttggagga ggatggccca gcttctgata gcacagttga aagtagtgat 4140
gtctatgaaa ctataattag tattgatgat aaagggcagg ccatgtacag ttttggtcga 4200
tttgactcct ccataataag aataaagaac cctgaagatg gtgagttgat agaccagtct 4260
gaagagggct tgatagcaac gggagtgaga attagtgagc tgcccttgaa agactgtgct 4320
caaggtgtga aaaagaagaa atctgagggc agttccattg gtgagtctac acgaattcgc 4380
tgtgatgatt gtggcttctt agcagatgga ctgagtggac tgaatgttca catagccatg 4440
aagcatccta caaaagagaa gcacttccat tgtttactct gtggaaagtc gttctatacc 4500
gaaagcaacc ttcaccagca tctggctagt gccggccaca tgagaaatga gcaggccagt 4560
gtggaggagc ttccggaggg aggggccacc tttaaatgtg tcaagtgcac agagcccttt 4620
gattctgaac agaatttatt tctacatatt aaaggacagc atgaggaatt gctgcgggag 4680
gtgaataagt atatagtgga agacactgag caaatcaacc gcgagaggga ggaaaaccag 4740
ggaaacgtct gcaagtattg tgggaagatg tgtcgaagca gcaactcgat ggccttcctg 4800
gcacacattc gcactcacac aggatcaaaa ccattcaagt gcaagatatg ccattttgca 4860
acagctcagc ttggagatgc cagaaaccat gtgaaaaggc accttgggat gagggaatac 4920
aagtgtcatg tctgtggggt tgcttttgta atgaagaagc acttaaatac tcatctacta 4980
ggcaagcatg gagttggcac cccaaaagaa aggaaattta catgccactt atgtgataga 5040
agtttcacag agaagtgggc cctgaacaac cacatgaaac tccacacggg agaaaagccg 5100
tttaaatgta cctggcccac gtgccattac tcattcctca cagcctccgc aatgaaagac 5160
cactacagga cgcacacagg cgagaagtcg tttctgtgtg acctctgcgg ctttgccggc 5220
gggacccgcc acgccctcac caagcatcgc agacagcaca caggagaaaa acctttcaag 5280
tgcgatgagt gtaactttgc ctccacaact cagtcccatt tgactcggca taaacgtgtc 5340
cacactggag aaaagcccta cagatgcccc tggtgtgact acaggtcaaa ctgtgctgaa 5400
aatatccgca aacacattct gcatactggc aaacatgaag gagtcaagat gtacaactgt 5460

cccaagtgtg actacgggac caacgtcccg gtggagttcc ggaaccattt gaaggaacag 5520
catcctgaca tcgaaaaccc ggacctcgct tacctgcatg ctggcattgt ttccaagtcg 5580
tacgagtgcc gtctaaaggg acaaggagcc accttcgtgg agacagacag ccccttcacc 5640
gcggcggcct tggcagaaga gcccctcgtc aaggagaagc ccctcagaag cagcaggagg 5700
ccagcgccgc cccctgagca ggtgcagcag gtcatcatct tccagggcta cgacggggag 5760
tttgccctgg acccctcggt ggaggagacg gccgccgcca cgctgcagac gctggccatg 5820
gccggccagg tggcccgggt ggtgcatatc acggaggatg gccaggtcat cgccacgagt 5880
cagagcgggg cacatgtagg cagcgtggtg cccggaccca tcctccccga gcagctggct 5940
gatggagcca cccaggtggt cgtcgtgggg ggctccatgg aaggccacgg catggatgag 6000
tccctcagtc caggtggcgc tgtgatacaa caggtgacca agcaggagat tttaaacctc 6060
tcggaggctg gagtcgctcc ccccgaggca tcctcagccc tggatgcatt gctctgtgcg 6120
gtcactgaat taggggaggt ggagggcagg gctgggctcg aggagcaagg caggcccggc 6180
gccaaagacg tgctgatcca gctgcccggg caggaggtct cccatgtggc tgccgacccc 6240
gaggcccccg agatccagat gttcccacag gcccaggaga gcccggccgc cgtggaggtg 6300
ctcacccagg tggtccatcc ctcagcagcc atggcctctc aggagcgggc acaggtggcc 6360
ttcaagaaga tggtccaggg cgtcctccag tttgctgtgt gtgacacggc cgcggccggc 6420
cagttggtca aggacggtgt cacccaggtg gtggtgagcg aagagggtgc cgtccacatg 6480
gtcgccgggg agggtgccca gatcatcatg caggaggcgc agggcgagca catggatctg 6540
gtggagtccg acggggagat ctcgcagatc atcgtgacag aggagctggt ccaggccatg 6600
gtgcaggagt ccagtggcgg cttctccgag ggcaccacgc actacatcct gacagagctg 6660
cccccagggg tgcaggacga gccgggcctg tactcccaca ccgtgctgga gactgcggac 6720
tcgcaggaac tcctgcaggc cggggccacg ctaggcacag aggccggggc cccaagcagg 6780
gcagagcagc tggccagcgt ggtcatctac acccaggagg gctcctcggc cgcggcggca 6840
attcagagcc aaagagaaag cagcgaactc caggaagcat gagacgcgcg gcacctttac 6900
tcagcacagg gcaggtgtgg gaaggtccag cttcggtggg ggaccgtgtt ccctgagctt 6960
catctgaaac cttcaaaacc atgaggacaa ggctcccgtg agctctgagc atgccctccc 7020
agcgagagtc acactggcca ccagccaggc gcccacagag ggtaccgtgg gctgggcctc 7080
ggggagcagg ctgccaagtg caggggaggg ccgggcgcag gccgcacagg gagctccggt 7140
ccactggggg cccttcgatc agtggcctcc cgcttcgtcc tggccgctgt gctgaagaga 7200
agccaagtgt tgttggtgtt tttctctccc aagtgttttc ccatttcagt tatcagaagg 7260
tcatggccgt ggggaaagtg gtgaagatac ccctcctggc ttggggtgca cctgcttgtg 7320
cagtcagcat gtagctgcct ttccatttca ttctctactg ggctaaaaat tgcagctaca 7380
agtgttacca tcttgaagca gtccacttcc attcaatttt tttttttaa ttttagaata 7440
acagtgtccc cataccaaag gaagcctgct agctcatttc atgtataaat ttcccatctt 7500
caaacagttt aggtgtattt gttgctctgg tcacattctg cataaaagaa atcctcttaa 7560
gcctatggtt aagaaaagcc ttgaagttta tattcagtta aaatatatgt cggtggagat 7620
agccagtgct tctaattttg acttagtttc atacagtaaa gcctaaatgt gaaacgcaca 7680
cgctggaaga tattgttcct atcaatattt tgcttttat aacaagggtt tgttcatatt 7740
gatgccattt ttgcaggatt tcttcgtgat ttctgtccat atgaaaatgc tgacattaaa 7800
cattaacaca tggagaccgt gccctgtggc cctgccgtgg ctgccagcat ggtctgtgtt 7860
tccttgtgga ttcacctgtg gccctgctgt ggccaccagc atggtctgtg tcctcgtgga 7920
ttcactgcag ctgtcggatg cgagtttctg tcataatcat ttgtttcctg atacaattgt 7980
tcttattctt ttccaaaact gtaaaataat ctcctccctc aaatgcaaag gttgtttttg 8040
ttctgtttct gttttctttg aaataaaatt ataacgttaa aagataaaaa aaaaaaaaaa 8100
aaaaactgcg tgtc 8114

<210> 68
<211> 1530
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2801633CB1

<400> 68
ggcgcggctt ttgcttgtag ctccagccag agctcggtta gggcctcatc gctctgctcc 60
cgctccttag ggaagcctcg gtgattctgc cacagcctca gcctctgtgg ctctgtgacc 120
tgccggtatt ggatgattcg tatctaagac tctgggacac tcctgaagtc gggaaatgga 180
actcttaaca ttcaaggatg tggccataga attctctcca gaagagtgga aatgtctgga 240
catttcccag cagaatttat atagagatgt gatgttggag aactacagaa acctggtctc 300
cctgggtgtt actatctcta acccagacct ggtcaccagt ctggagcaaa gaaaagagcc 360
ctacaatttg aagatacatg aaacagcagc cagaccccca gctgtgtgtt ctcatttcac 420

ccaaaacctt tggacagtgc agggcataga agattcattc cacaaactta taccaaaagg 480
acatgagaaa cgtggacatg agaatttaag aaaaacttgt aaaagtataa atgagtgtaa 540
ggtgcagaaa ggtggttata atagaattaa ccaatgctta ttaactaccc agaaaaaaac 600
aattcaatct aatatatgtg tcaaagtttt tcataaattt tcaaattcaa acaaagataa 660
gataagatat actggagata aaacctttaa atgtaaagaa tgtggcaaat catttcacgt 720
gctctcacgc ctaactcaac acaaaagaat tcatactgga gagaacccct acacatgtga 780
agaatgtggc aaagccttta attggtcctc aattcttact aaacataaga gaattcatgc 840
cagagagaaa ttctacaagt gtgaagaatg tggtaaaggc tttactcggt cctcacacct 900
tactaaacat aagagaattc atactggaga gaaactctac acatgaaaaa attgacaaag 960
cttttaaccg caactcaatc tgttctaaac ataagagaaa tggtattggt gagaagccat 1020
aaaaatatga aaaatgtgga aaacccttca aatgcttgtc acatcttact gaatataatt 1080
cttactgcag aaaacccctа ggaatattaa aagtgtggca aaacctttaa ccaatgctca 1140
tatctctttg cacatgatag catttatact tgagaaaaat tgtacaaata tagaaaatgt 1200
agaaaagcca ttaatgccta ctcatgtgtt actaaatatc agagagtttg tacttaataa 1260
aaggattata aatgtagtat ttgttgaaag acctattaga aaatacaggt cttaaaagtg 1320
aagaagagta ttctgaagat agacaataga aatagtaaga gggttgtagt acctgtactt 1380
gcatcatggg tcttattgtg catatttcat actagaagaa aaccctgaag cagttgccca 1440
aactttcttg acattagaga atttatatcg gaaagaaatt ttacaaatgt gataaatttg 1500
gaaaagcaaa aacaaaaaac aaaaaaaaaa 1530

<210> 69
<211> 2026
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7493525CB1

<400> 69
gctccaggtc tccccttcgc tgctctgtgt cctctgctcc tagaggccca acatctgtgg 60
ccctgtgacc tgcaggtatt gggagaccca cagctaagac accgggaccc cctgaaagcc 120
tagaaatgga cgacttgaaa tatggagtgt atcctctcaa ggaagcaagt ggatgccctg 180
gggctgagag gaatcttcta gtttactctt attttgaaaa ggagacattg acatttaggg 240
atgtggccat agaattctct ctggaggagt gggaatgcct gaaccctgct cagcagaatt 300
tatatatgaa tgtgatgtta gaaaactaca aaaacctggt cttcttggca ggtgttgctg 360
tctctaagca agacccagtc acctgtctgg agcaagaaaa agagccctgg aatatgaaga 420
gacatgagat ggtggatgaa cccccagcta tgtgttctta ttttaccaaa gacctttggc 480
cagagcaaga cataaaagat tcttttcaac aagtaatact gagaagatat ggcaaatgtg 540
aacatgagaa tttacagtta agaaaaggct ccgcaagtgt agatgagtat aaggtgcaca 600
aagaaggtta taatgagcta aaccagtgtt tgacaactac ccagagcaaa atatttccat 660
gtgataaata tgtgaaagtc tttcataaat ttttaaatgc aaatagacac aagacaagac 720
atactggaga gaaacctttc aaatgtaaaa aatgtgatga atcattttgc atgcttttac 780
acctaagtca acataaaaga attcatatta gagagaattc ttaccaatgt gaagaatgtg 840
gcaaagcttt taaatggttc tcaaccctta ctagacacaa gagaattcat actggagaga 900
aacccttcaa atgtgaagaa tgtggcaaag cttttaagca ctcctcaacc cttactacac 960
ataagatgat tcatactgga gagaaaccct acagatgtga agaatgtggc aaagccttct 1020
accattcttc acaccttact acacataagg taattcatac tggagagaag cccttcaaat 1080
gtgaagaatg tggtaaagct tttaaccacc cttcagccct tactacacat aagttcattc 1140
atgttaaaga aaaaccctac aaatgtgaag aatgtgacaa agcttttaac cgattctcat 1200
accttactaa acataagata attcattctg gagagaaatc ttacaaatgt gaacaatgtg 1260
gcaaaggctt taactggtct tcaacccttа caaaacatag aagaattcat actggagaga 1320
aacccctacaa atgtgaagaa tgtggcaaag cctttaatgt gtcttcacac cttactacac 1380
ataagatgat tcatactgga gagaaaccct acaaatgtga agaatgtggc aaagccttta 1440
accactcctc aaaacttact atacataaga taattcatac tggagagaaa ccttacaaat 1500
gtgaagaatg tggcaaagct tttaaccaat cctcaaacct taccaaacat aagataattc 1560
atactggaga gaaactctac aaatgtgaag aatgtggcaa agcttttaac cgatcctcaa 1620
accttactac acataagaga attcacactg gagagaaacc ctacaaatgt gaagaatgtg 1680
gcaaagcttt caaccgatcc tcaaacctta ctaaacataa cataattcat actggagaga 1740
aatcttacaa atgtgaagaa tgtggtaaag cctttaacca atcctcaact cttactaaac 1800
ataggaaaat tcagcagggc atggtggctc atgcctgtaa tcccaacact ttgagaggac 1860
taggtgagca gatcgcgagg tcaggagttc aagaccagcc tggccaacat ggtaaaaccc 1920
catctctact aaaaatacaa aaatttgctg ggtgtggtgg caggcgcctg taatcccagc 1980

tacttggctc actgcaggct ccgccttccg ggttcacgcc attctc 2026

<210> 70
<211> 1724
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7021892CB1

<400> 70
atgctgttgg tggccaccat attggacagc ccgggtctag acccttttta ccacttcaac 60
tctatgtcga ataatgaacg atattcttca ttccagctag catgtgaaat tcagaagtca 120
gcagccatga ctcttcacgt ctgtacccgt attgcctggt acaagggata tcacatagta 180
ggtaaaaacc tgtccaatag taacaatctc aatgatggca gaatgaaatc agaatctgat 240
tggattaaaa aggagggaaa gggtgtggct aaggtgggtg gagacaccct ctggtataaa 300
agcccctggc aggcagctct cactccagat ctgagctgtc cgcagaagca gctggaggcc 360
aggggagaaa ctccagaagg agagacattt gccatggctg agcacttcaa acagatcatt 420
agatgtcctg tctgtctaaa agatcttgaa gaagccgtgc aactgaaatg tggatatgcc 480
tgctgcctcc agtgcctcaa ttcactccag aaggagcccg atggggaagg tttactgtgc 540
cgtttctgct ctgtggtctc tcagaaggat gacatcaagc ccaagtacaa gctgagggcg 600
ctggtttcca tcatcaagga actagagccc aagctgaaat ctgttctaac aatgaaccca 660
aggatgagga agtttcaagt ggatatgacg ttcgatgtgg acacagccaa caactatctc 720
atcatttctg aagacctgag gagtttccga agtggggatt tgagccagaa taggaaggag 780
caagctgaga ggttcgacac tgccctgtgc gtcctgggca cccctcgctt cacttccggc 840
cgccattact gggaggtgga cgtgggcacc agccaagtgt gggatgtggg cgtgtgcaag 900
gaatctgtca accgacaggg gaagattgtg ctttcttcag aacacggctt cttgactgtg 960
ggttgcagag aaggaaaggt ctttgctgcc agcactgtgc ctatgactcc tctctgggtg 1020
agtccccagt tgcacagagt ggggattttc ctggatgtag gtatgaggtc cattgccttt 1080
tacaatgtta gtgatgggtg ccatatctac acattcatcg agattcctgt ttgcgagccc 1140
tggcgtccat tttttgctca taaacgtgga agtcaagatg atcagagcat cctgagtatc 1200
tgttctgtga tcaatccatc cgctgccagt gccccagttt cttctgaggg aaagtaaata 1260
aacatttgaa cataatcatc tttaggaagt ttcagtgctc ccatagccat agctaagaac 1320
ttttccgcta gatacacata ggtacaaagg gacagaaggg aaagagccat cactctgtaa 1380
accatgaaca gaaagcaatt atattaatga ggggaaaatt acattgtact taaagttgtc 1440
atgtcgtttt ctctgggctt atgttatatt tctgttcata aatatttgca aaaaaaacaa 1500
aaggggcgcg cgattatgag ctcgtcgacc gggattatcc ggccggtact ggggcgtaca 1560
gtttcctatg gggggcggat tagacggcgt aacaggctag ctgtcccggt gaatgtccgc 1620
caccacccat acagcacaca ccttggaagg ggggccaaac gcgcacacga aagcgaggcg 1680
cggagccgcg acaccgctgg cgaagcaggc acgcacacca acgg 1724